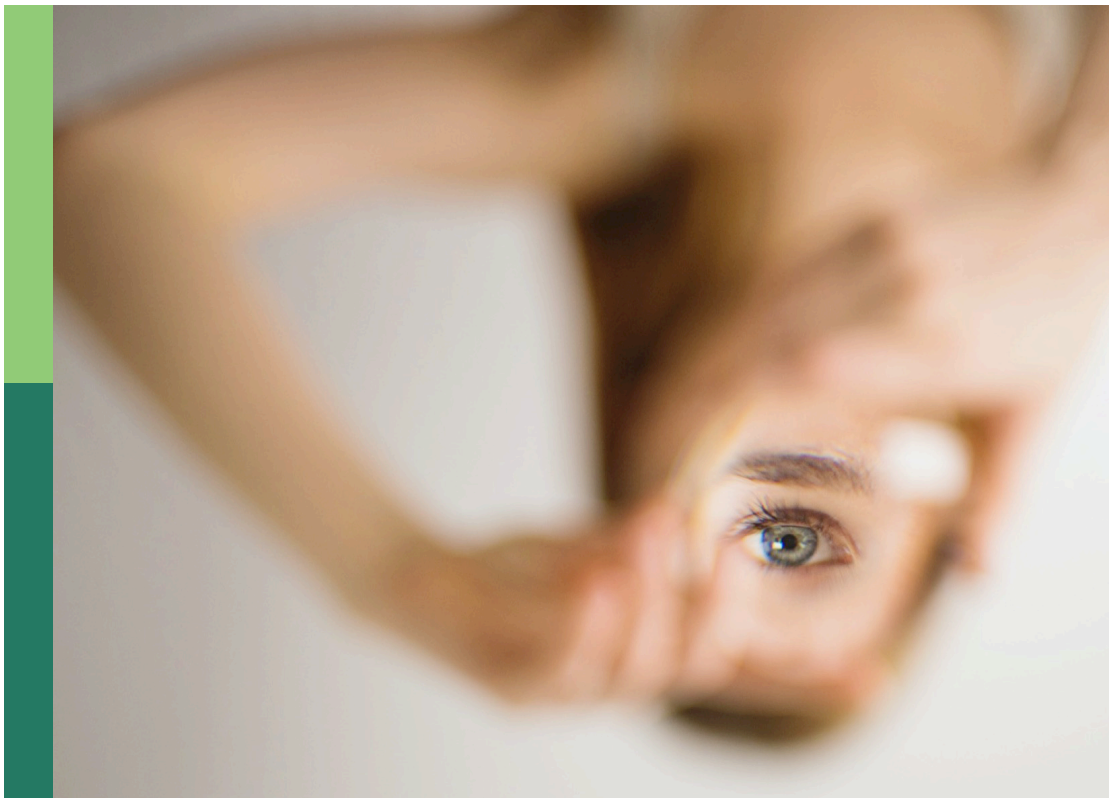
frontiers | Research Topics

From “modern” to “postmodern” psychology: Is there a way past?

Edited by
Barbara Hanfstingl, Timo Gnambs, Peter Adriaan Edelsbrunner, Jana Uher and Ulrich Dettweiler

Published in
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-83251-944-8
DOI 10.3389/978-2-83251-944-8

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

From “modern” to “postmodern” psychology: Is there a way past?

Topic editors

Barbara Hanfstingl — University of Klagenfurt, Austria
Timo Gnambs — Leibniz Institute for Educational Trajectories (LG), Germany
Peter Adriaan Edelsbrunner — ETH Zürich, Switzerland
Jana Uher — University of Greenwich, United Kingdom
Ulrich Dettweiler — University of Stavanger, Norway

Citation

Hanfstingl, B., Gnambs, T., Edelsbrunner, P. A., Uher, J., Dettweiler, U., eds. (2023). *From “modern” to “postmodern” psychology: Is there a way past?* Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83251-944-8

Table of contents

frontiers | Frontiers in Psychology

TYPE Editorial
PUBLISHED 03 March 2023
DOI 10.3389/fpsyg.2023.1091721

OPEN ACCESS

EDITED AND REVIEWED BY
Anna M. Borghi,
Sapienza University of Rome, Italy

*CORRESPONDENCE
Barbara Hanfstingl
✉ barbara.hanfstingl@aau.at

SPECIALTY SECTION
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

RECEIVED 07 November 2022
ACCEPTED 07 February 2023
PUBLISHED 03 March 2023

CITATION
Hanfstingl B, Uher J, Edelsbrunner PA, Dettweiler U and Gnambs T (2023) Editorial: From "modern" to "postmodern" psychology: Is there a way past? *Front. Psychol.* 14:1091721. doi: 10.3389/fpsyg.2023.1091721

Editorial: From "modern" to "postmodern" psychology: Is there a way past?

Barbara Hanfstingl[1]*, Jana Uher[2], Peter A. Edelsbrunner[3], Ulrich Dettweiler[4] and Timo Gnambs[5]

[1]Institute for School and Instructional Development, University of Klagenfurt, Klagenfurt, Austria, [2]School of Human Sciences, University of Greenwich, London, United Kingdom, [3]Department of Humanities, Political and Social Sciences, ETH Zürich, Zürich, Switzerland, [4]Faculty of Arts and Education, University of Stavanger, Stavanger, Norway, [5]Educational Measurement, Leibniz Institute for Educational Trajectories (LIfBi), Bamberg, Germany

KEYWORDS

modern psychology, postmodern psychology, methodology, psychology of science, epistemology in psychology, psychological phenomena and processes

Editorial on the Research Topic
From "modern" to "postmodern" psychology: Is there a way past?

Contemporary psychology is facing profound problems and various obstacles to advancing its research, as reflected in its continued crises in replicability, confidence, generalizability, and validity. "Modern" paradigms–involving beliefs in determinative cause-effect relations between the elements of an objectively given world, which are thus amenable to experimental, rational exploration and mostly linear statistical analyses–often no longer do justice to the complexity of psychology's contemporary research questions. Critical analyses of established concepts and approaches have not yet been sufficiently considered in mainstream theorizing nor have adequate consequences been drawn from them to advance our understanding of the phenomena of mind and behavior as well as to elaborate overarching frameworks and to further methodologies and methods that are suited for their exploration.

This Research Topic assembled contributions from authors with expertise in different specialties to work toward developing a new understanding of psychological science, aimed at tackling current problems and devising possible solutions by exploring the promises of "postmodernism" as well as of further epistemologies and research paradigms beyond. In the context of science, "postmodernism" has no overarching meaning. It is associated with epistemological developments after Karl Popper's critical rationalism, such as constructivism, systemic approaches, and epistemological as well as a methodological plurality. To avoid fruitless doctrinal dispute, we did not insist on the terms "modern" and "postmodern" nor on any narrow definition of them. Instead, we invited papers proposing new ideas and solutions that may have the potential to tackle the epistemological, conceptual, and methodological challenges of "modern" psychology and to improve research quality through more critical and more in-depth reconsiderations than commonly done in currently popular calls for "robust analysis," preregistration, replicability, and open science. Our key questions were to what extent we need to abandon the ideas of critical rationalism, to what extent we need to integrate concepts and methodological strategies from other disciplines, and to what extent we should focus on entirely new problem-solving strategies.

The current Research Topic includes 15 articles from different world regions that have discussed these issues and key questions from multiple perspectives. Here, we briefly summarize these contributions to highlight their diversity as well as central themes in their future-oriented reflections and proposals for solutions.

Holtz contrasts postmodernist with critical rationalist conceptualizations and analyses the main differences. He shows that, rather unexpectedly, Karl Popper's and David Deutsch's understanding of objective knowledge, progress, and methods is not that different from Kenneth Gergen's understanding. All of them agree that focusing only on a specific scientific method neither justifies nor validates psychological knowledge. Popper and Deutsch see scientific progress in the formulation of "better" theories, which are derived from formal and logical reasoning, whereas Gergen sees scientific progress much more in its abilities to address real-world problems in the context of culture and society. Holtz argues for a joined next step for developing epistemology in psychological science.

In a similar vein, Mazur revives early criticisms of positivism that has been voiced in pre-postmodernist times already. He suggests that, rather than addressing the shortcomings of positivistic epistemology by means of postmodernism, psychology would be better served by the deeper, more consequential reflections of Sapientia, a form of metaphysical wisdom that asserts the power of science as a method, while also critically and cautiously supporting the polyvalence and complexities of life as highlighted in postmodern thought.

Veraksa et al. propose dialectical thinking as a basis for developing psychology from a modern to a postmodern science. Dialectical thinking recognizes the importance of contradiction, change and synthesis. This includes recognizing the values as well as the limitations of modern epistemological approaches, such as both universalistic formal analysis (often associated with modernist approaches) and relativistic analysis (often associated with the rejection of modernist approaches). Veraksa et al. present dialectic thinking as a powerful processual approach to conceiving scientific thinking and advancing the development of scientific knowledge.

Iso-Ahola points out that the scientific truth that we aspire has to be seen generally as time-related, context-related, and method-related. A successful replication does not automatically mean validation of findings when simple measurement problems, like the reliability of a scale, remain unconsidered. He further points to the influence of methodical artifacts, stability, temporality, context-dependence, and the implicity of many psychological phenomena, which all disturb the accuracy of psychological constructs. Therefore, Iso-Ahola suggests focusing on psychological phenomena in replication studies and evaluating them primarily on a theoretical level rather than only on a methodical level.

Krueger highlights the mutual enrichment that is needed between forward-looking experimental psychology and backward-looking historical psychology in a postmodern scientific era, given that prediction and explanation make no sense without the other. In other words, they can be seen as the same, only the direction of the time flow is reversed. Krueger supports his argument with Bayesian considerations and the diagnostic ratio, showing that probability and effectiveness are inversely related so that rare causes with high effectiveness must also be considered in psychological explanatory models. He supports this argument with thought experiments based on the three historical case studies of Philipp von Hutten, Gonzalo Guerrero, and Robinson Crusoe.

Another fundamental statement comes from Rabeyron. He uncovers problems of psychological mainstream methodology with particular contents in psychology that are associated with findings that have led to the Nobel Prize in physics 2022 for research in quantum entanglement by Alain Aspect, John Clauser, and Anton Zeilinger. Based on two examples of psi research, the Ganzfeld experiment and the Bem experiment, Rabeyron argues that much developmental work on methodology and theory has to be done to explain psychological phenomena that are as complex as psi is assumed to be. Rabeyron connects his considerations with earlier work that is explicitly associated with physics, for example, Lucadou's Model of Pragmatic Information.

Focusing on knowledge generation in psychological experiments, Mayrhofer et al. analyze their underlying philosophy of science. They state that researchers must reduce and pre-structure the phenomena of interest in order to (re)create narrowly defined phenomena in a controlled environment and develop meaningful research questions and hypotheses. That is, rather than a copy of "reality," the experimental setup is an active construction by the researchers and reflects their pre-experimental understanding. Mayrhofer et al. demonstrate that postmodern concepts have always been at the heart of psychological experiments and can therefore be fruitfully applied to sharpen the theoretical and empirical basis of experiments.

From a more societal perspective, Guyon highlights the tension that exists between the scientific imperative of quantification in experimental psychology and the social imperative of its actual use and implementation in psychology. Specifically, standardization, control and regulation are meant to provide scientifically validated findings that serve to support public decision-making. But ultimately, results depend on scientists' subjective choices (e.g., of statistical models, and interpretations) and can be apprehended only through the prism of social practice.

Emphasizing the need for more theory, Burghardt and Bodansky note that psychology as a science has left the first phase of exploratory research in favor of theory-driven research. To manage this transition, Burghardt and Bodansky present five key challenges. Challenge One is about how to best support researchers to advance the field. Challenge Two concerns psychology's transition from protoscience to paradigmatic science, in which scientists are challenged to develop robust paradigms that help associate and restructure currently unrelated findings and theories. Challenge Three involves a revised methodology needed along the lines of Lakatos, who developed Popper's critical rationalism into a more theory-friendly research advance. Challenge Four stresses the need for harmonizing processes between theory and evidence, in line with Klaus Holzkamp's ideas. Finally, Burghardt and Bodansky present as Challenge Five a 10-point checklist for good research.

Does psychological science have not only a neglected relationship to theory but also a peculiar relationship to methods that are worthy of scrutiny? Several authors agree. For example, Mayrhofer and Hutmacher report on Gerd Jüttemann's so-called "principle of inversion," which has befallen psychology as a science, by allowing its content to be dictated by methodology. This means a contrast between strict methodological requirements and the comprehensive and often unclear thematic content of psychological research. In other sciences, by contrast, the content is in the foreground. As a consequence, psychological science must abandon the notion of a hierarchy of power in methods, with just quantitative ones at the top. Instead, postmodern science argues for focusing on psychological phenomena that can be understood only through the application of a plurality of methods.

Borgstede and Stolz discuss the different importance of replication in deductive, variable-based, more quantitatively oriented research (top-down generalization) vs. in inductive, case-based, more qualitatively oriented research (bottom-up generalization). If replication fails in the former, the theory would have to be falsified because the approach assumes the need for the generalizability of theories. In less formalized inductive research, in contrast, a failed replication leads to considerations about the limitations of a theory's validity. Borgstede and Stolz argue for a more frequent and open-minded use of bottom-up generalization because statistical sample-based generalization modeling is—from the perspective of formal logic—unattainable in social and psychological science.

Edelsbrunner discusses various statistical and conceptual rationales for generating sum scores across items (e.g., in psychological tests), arguing that any given score can only represent the particular theoretical model that has been used to create it. Therefore, he demands that researchers explain why and how they want to justify the use of scores either through a specific theoretical rationale (e.g., conceptualization

TABLE 1 Comparison of key assumptions of "modern" vs. "postmodern" and further epistemologies in psychology discussed in this Research Topic.

"Modern" psychology	"Postmodern" psychology and beyond	Discussed by
Protoscience	Paradigmatic science	Burghardt and Bodansky
Orientation toward traditional natural sciences	The necessary focus on psychological phenomena and their peculiarities, requiring the involvement of sciences beyond the traditional natural sciences	Burghardt and Bodansky; Guyon; Iso-Ahola; Mayrhofer and Hutmacher; Mayrhofer et al.; Mazur; Rabeyron; Uher
Knowledge generation through authority and scientific hierarchy, focusing on important scientist personalities ("VIPs of science")	Collective knowledge generation with less focus on influential individual researchers but instead on diversity in researchers and their sociocultural and research backgrounds	Burghardt and Bodansky; Guyon; Hanfstingl; Mayrhofer et al.; Mazur; Uher; Veraksa et al.; Zitzmann and Loreth; conversely discussed by Holtz
Generalized theories are regarded as valid across contexts and populations	Theories and approaches that are valid only locally or temporarily, with this limited validity being regarded as a strength rather than a weakness; Accepting the contextuality of findings	Borgstede and Scholz; Guyon; Hanfstingl; Holtz; Mayrhofer and Hutmacher; Mayrhofer et al.; Iso-Ahola; Mazur; Uher; Veraksa et al.
Knowledge from single studies	Knowledge from meta-analyses, meta-syntheses, and reviews	Hanfstingl; Rabeyron; Uher
Primary focus on empirical studies	Focus on both theory development and empirical studies	Borgstede and Scholz; Burghardt and Bodansky; Edelsbrunner; Guyon; Hanfstingl; Iso-Ahola; Mayrhofer and Hutmacher; Mayrhofer et al.; Mazur; Uher; Veraksa et al.; Zitzmann and Loreth
Implicit hierarchy of the quality of scientific methods with quantitative methods at the top	Plausibility of a method's applicability and its appropriateness to the peculiarities of the study phenomena	Borgstede and Scholz; Guyon; Holtz; Krueger; Mayrhofer and Hutmacher; Mayrhofer et al.; Mazur; Uher; Rabeyron; Zitzmann and Loreth
Focus on only one method, one approach, or one theory	Manifold and complementary use of different methods, approaches, theories or even disciplines to gain new insights	Borgstede and Scholz; Hanfstingl; Holtz; Krueger; Mayrhofer and Hutmacher; Iso-Ahola; Uher; Veraksa et al.; Zitzmann and Loreth
Orientation toward psychological constructs	Orientation toward psychological phenomena in themselves away from beliefs about them as reflected in everyday constructs	Borgstede and Scholz; Burghardt and Bodansky; Edelsbrunner; Guyon; Hanfstingl; Iso-Ahola; Mayrhofer and Hutmacher; Mayrhofer et al.; Mazur; Uher; Veraksa et al.
Replicated studies produce valid knowledge, unreplicated studies produce invalid knowledge	Re-interpretation of replication as a method for examining generalizability and contextuality	Borgstede and Scholz; Iso-Ahola; Mayrhofer et al.; Mazur; Rabeyron; Uher
Statistics as a truth-generator	Re-interpretation of statistics as socio-constructivist activity legitimately dependent on the researcher	Borgstede and Scholz; Burghardt and Bodansky; Edelsbrunner; Guyon; Hanfstingl; Holtz; Iso-Ahola; Mayrhofer and Hutmacher; Mazur; Uher; Zitzmann and Loreth
Rules on how to use methodological principles	Arbitrariness in the use of methodological principles	Holtz; Krueger; Zitzmann and Loreth; conversely discussed by Mayrhofer et al.; Uher
Accepting knowledge as valid because it has been published in peer-reviewed journals	A critical look at the processes of knowledge generation and their transparency	Guyon; Rabeyron

or definition) or a specific statistical rationale (e.g., through a statistical model). The general aim of Edelsbrunner is to get beyond long-standing consensus views, such as that traditional latent variable models could be adequate representations underlying any measurement process.

Zitzmann and Loreth critically discuss how researchers' preferences of statistical methods often influence their mutual approval as researchers (e.g., as "being" Frequentist vs. Bayesian) and how this hampers knowledge dissemination (e.g., through overly critical reviews)—and thus scientific progress. They argue for strengthening researchers' shared identity as psychologists (e.g., by facilitating non-mainstream publications in the same respected journals) without having to give up their disapproval of lower levels of identity (e.g., preferences for particular methods). In particular, mutual tolerance and respect of others as equals enable much-needed critical discussion and serious debate.

With an even stronger focus on methodological research practices, Uher analyzes in-depth the philosophical-theoretical foundations on which rating scales are built, revealing a dense network of 12 complexes of problematic concepts, misconceived assumptions, and fallacies. Uher demonstrates how—through the popularity of rating scales and their uncritical interpretation as enabling psychological "measurement"—these problems have become institutionalized in a wide range of psychological practices, thereby perpetuating psychology's crises (e.g., replication, validity, generalizability, and confidence). To tackle these problems holistically, Uher derives from all 12 problem complexes specific theoretical concepts, methodologies, and methods as well as key directions of development, highlighting the necessity to explore individuals as complex living beings and to consider the study phenomena's peculiarities (e.g., momentariness, contextuality, and intra-individual variation) as well as the inherent anthropogenicity of any research on individuals.

Hanfstingl points out that methodology has to develop a new understanding of objectivity to meet future requirements of validity in psychological science. Taking up the argument from feministic and sociological research that different perspectives must be considered to enhance objectivity, Hanfstingl holds that different theories and methods are to be defined as concrete perspectives on a psychological phenomenon. Psychological research methodology has many methodical tools available to systematically apply different perspectives on a phenomenon, such as, for example, specification analyses, meta-analyses, combinatorial meta-analyses, or approaches combining any of those. Hanfstingl argues that research programs should be designed on the base of these formally explicated perspectives and around a psychological phenomenon of interest.

All of these contributions highlight the need for a paradigm shift in psychological science. This shift is thereby not seen as a complete break from what has been done so far but rather as a gradual change developed and implemented over generations of researchers. A pivotal common thread is a need for a greater focus on the integration of new developments in theory, methodology, and methods in order to meet the requirements of future research and contemporary real-world problems. A key insight is that psychological knowledge is much more complex than the mainstream understanding currently represents. Many psychologists' understanding of validity and objectivity is deeply shaken because they are struggling to find consensus even regarding the meaning of "success" in the replication of empirical studies. Consequently, there is a substantial need for new epistemological and methodological developments, in particular, because these developments have not yet reached mainstream research and its ongoing debates. The old ways of doing research often do not work anymore, whereas new developments are not yet fully elaborated and functional.

All the authors of our Research Topic are working toward solutions and they have contributed ideas and strategies for dealing with these new insights that can no longer be ignored in psychological science. These insights call for a closer look at what "modern" science in psychology already offers, at the new directions that "post-modern" and other epistemologies have opened up, and at how other disciplines are already dealing successfully with this change. Many authors also note that psychology has evolved from a protoscience to a paradigmatic science with all the consequences that this entails, at both the theoretical and the methodological level. They call for intensified research on theory in psychology and much greater use of theoretical knowledge to gain new insights. Many authors of our Research Topic ascertained that focusing on the psychological phenomena in themselves rather than just on psychological constructs about them is more helpful in gaining new insights. They offer new variants and alternatives for generating scientific data and interpreting results that meet the complexity of psychological study phenomena more appropriately and that enable the generation of psychological findings with higher validity, replicability, generalizability, and confidence.

To outline such a paradigm shift, we believe that today's understanding of science has changed in several ways compared to the understanding of science by earlier generations of scientists. In Table 1, we summarize the main differences between the traditional "modern" view of psychological science and the recent "postmodern" and further developments of psychological science including some controversial ideas discussed in the articles of our Research Topic without any claim of completeness.

It is our hope that this compilation of research papers will contribute new ideas, theories, concepts, and methodologies to current debates on psychological research practices and will provide good food for thought to help psychologist tackle their current challenges and advance their discipline and its research meaningfully.

Author contributions

BH and JU developed the idea for this article, drafted, and finalized the manuscript. PE, UD, and TG provided feedback and suggestions. All authors approved the manuscript for submission.

Acknowledgments

We would like to thank all authors who have contributed with their ideas to the Research Topic in its present form as well as

all those who, as editors and reviewers, have contributed to a significant increase in quality.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

frontiers in Psychology

REVIEW
published: 16 September 2020
doi: 10.3389/fpsyg.2020.02183

Replication and the Establishment of Scientific Truth

*Seppo E. Iso-Ahola**

Department of Kinesiology, School of Public Health, University of Maryland, College Park, College Park, MD, United States

The idea of replication is based on the premise that there are empirical regularities or universal laws to be replicated and verified, and the scientific method is adequate for doing it. Scientific truth, however, is not absolute but relative to time, context, and the method used. Time and context are inextricably intertwined in that time (e.g., Christmas Day vs. New Year's Day) creates different contexts for behaviors and contexts create different experiences of time, rendering psychological phenomena inherently variable. This means that internal and external conditions fluctuate and are different in a replication study vs. the original. Thus, a replication experiment is just another empirical investigation in an ongoing effort to establish scientific truth. Neither the original nor a replication is the final arbiter of whether or not something exists. Discovered patterns need not be permanent laws of human behavior proven by the pinpoint statistical verification through replication. To move forward, *phenomenon replications* are needed to investigate phenomena in different ways, forms, contexts, and times. Such investigations look at phenomena not just in terms the magnitude of their effects but also by their frequency, duration, and intensity in labs and real life. They will also shed light on the extent to which lab manipulations may make many phenomena subjectively conscious events and effects (e.g., causal attributions) when they are nonconsciously experienced in real life, or vice versa. As scientific knowledge in physics is temporary and incomplete, should it be any surprise that science can only provide "temporary winners" for psychological knowledge of human behavior?

Keywords: replication, reproducibility, scientific truth, scientific method, science, data

OPEN ACCESS

Edited by:
Peter Adriaan Edelsbrunner, ETH Zürich, Switzerland

Reviewed by:
David Trafimow, New Mexico State University, United States
Frank Zenker, Boğaziçi University, Turkey
Ayça Akan, Boğaziçi University, Turkey, in collaboration with reviewer FZ

***Correspondence:**
Seppo E. Iso-Ahola
isoahol@umd.edu

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *18 March 2020*
Accepted: *04 August 2020*
Published: *16 September 2020*

Citation:
Iso-Ahola SE (2020) Replication and the Establishment of Scientific Truth. Front. Psychol. 11:2183. doi: 10.3389/fpsyg.2020.02183

INTRODUCTION

This paper examines the nature of scientific and psychological truth and the role of replication in its establishment. It becomes evident from this examination that replication is only a part of the scientific method and does not have any special status in it. In fact, a so-called exact replication is just one type of replication or, at best, an approximation of the original study, and more generally, it is just another type of empirical study. "There are no critical tests of theories, and there are no objectively decisive replications" (Earp and Trafimow, 2015); no such thing as an exact or "direct" replication exists (Stroebe and Strack, 2014; Anderson et al., 2016; Rubin, 2019). Attempted exact replications cannot therefore become the final arbiters of truth any more than the original studies. In essence, then, every replication becomes a "constructive" (Lykken, 1968)

or "conceptual" (Crandall and Sherman, 2015) replication that may or may not add to the existing knowledge. Importantly, replications cannot provide yes-no answers to whether or not something exists, even though many argue that "direct replications test the basic existence of phenomena" (LeBel et al., 2017).

Exact or direct replications cannot be the final arbiters of scientific truth because (1) it is impossible to create conditions identical to the original test, and thus failures to replicate methods lead to failures to replicate results; (2) psychological phenomena are not limited to one specific form and condition, but are found in many different situations, and they are inherently subtle and variable due to the effects of time and context; (3) all effects are "interaction effects" even if laboratories are testing "main effects," leading to a difference between lab truth and real-life truth; and (4) the methods are psychometrically inadequate for giving categorical answers due to the problems of unreliability, invalidity, and sampling errors.

All attempts at exact replications are doomed to fail not because the investigated psychological phenomenon is not robust enough to reveal itself repeatedly, but because they attempt to replicate something whose existence is not limited to a specific context and specific time. By definition, exact replications rest on the assumption that a phenomenon exists only in the condition identical to the original, but overlook the reality that the tested phenomenon can exist in various forms and under different conditions. As the null hypothesis cannot be confirmed by non-existing exact replications, it is not surprising that such replication attempts have failed; logically, in fact, it is surprising that all *method replications* have not failed because they should have. Undoubtedly, some replications have failed because original studies were conducted under the "old rules" by not following today's stricter guidelines [e.g., p-hacking (Simmons et al., 2011)]. But focusing on past failures of method replications loses sight of the main thing: phenomenon and its boundary conditions.

In contrast, *phenomenon replications* test phenomena in varied forms, contexts, and times using different methods and consequently provide more nuanced and refined explanations than categorical declarations that the phenomenon is or is not real (Doyen et al., 2012; Carter et al., 2015; Gerber et al., 2016). They examine effects using multiple criteria other than the magnitude, such as frequency, duration, and intensity, as well as factors that give rise to phenomena and those that reduce their influence. They also shed light on boundary conditions and theories' strengths and weaknesses, thereby helping modify and expand theories. Phenomenon replications are constructive in nature but are not limited to single replications. Rather, they are programs of ongoing studies examining phenomena in different forms, contexts, and times.

Wegner (1994) presents good examples of phenomenon replications. He tested the same phenomenon (making an ironic error of mental control by actively avoiding thoughts of an object or action under mental load) employing different cognitive and behavioral tasks and contexts. Obviously, his methods and manipulations were different in different experimental situations, but they nevertheless reproduced the same result and additively and informatively showed that the effect is more pronounced in some tasks and situations (e.g., thought suppression) than others.

In a similar vein, Milgram's way of experimentally investigating obedience is just one of many ways of studying it. Although Milgram's findings have been replicated (Burger, 2009; Doliński et al., 2017), failures to directly replicate his original findings likely reflect methodological modifications, not necessarily that the phenomenon fails to influence outcomes (Elms, 2009). Moreover, experimental and non-experimental methods vary considerably in their sensitivity and ability to unveil phenomena. Since the same psychological phenomenon appears in different forms and to different degrees in varied contexts, a diversity of methodological approaches is necessary for replications of the originals findings. The accumulating, convergent evidence will provide a better understanding of possible false positive and false negative results of the original findings and, thus, of the nature of the investigated phenomenon.

Due to the fundamental limitation of the empirical method, no permanent scientific truth (or its absence) can be established by empirical replications. Phenomenon replications, however, can provide useful data and information and thereby improve estimates of the effects. But like original tests, replications can produce both false positives and false negatives, as these determinations are based on strict and arbitrary criteria, predominantly statistical thresholds (previously *p*-value, now Effect Size, Confidence Intervals, and Bayes Factor). A danger is that the search for underlying causes becomes largely a "statistical exercise" (Grice, 2014), even though "a statistical procedure is not an automatic, mechanical truth-generating machine for producing or verifying substantive causal theories" (Meehl, 1992, p. 143).

Statistical determination of scientific truth is based on the assumption that psychological attributes and phenomena are quantitative, but are they (Sherry, 2011; Grice, 2014)? If they are quantitative, a problem then becomes one of an agreement about the level at which psychological phenomena can be declared genuine and real. However, the agreement is not only about the quantity but more importantly, the meaning of numbers assumed to represent psychological constructs. This question of construct validity (Cronbach and Meehl, 1955) poses major challenges for psychological research in general and replications in particular, as "replicability does not equal validity" (Hussey and Hughes, 2020).

Taken together, however difficult or insurmountable the empirical testing of it would be, aliens' existence cannot be discounted, and so-called failed replications cannot declare any hypothesis logically or theoretically invalid. Similar to replications in physics, replications in psychology can only speak to observations about affect, cognition, and behavior in a specific context at a specific time. But this does not obviate the discovery of patterns that hold for certain situations and times (Lykken, 1991). In general, however, replications are logically tenable only if psychological phenomena can be claimed to be fixed and permanent entities, stable particles that can be described by absolute quantities. In the absence of well-founded claims, the basic premise of replication can be questioned.

NATURE OF SCIENTIFIC TRUTH

Implicit in the idea of empirical science is a question: What is truth anyway, and how is it determined? Replication studies attempt to answer this question by seeking to show if the original finding can be obtained again under similar experimental conditions. In science, truth, whatever its content, is said to exist to the extent that it is theoretically and empirically supported. Accordingly, a successful replication of an original finding is taken to mean that a truth exists, while a failure to replicate supposedly indicates the absence of the presumed truth. This fundamental axiom rests on the assumption that a scientific truth exists in the first place, and that it would reveal itself on researchers' empirical demands again and again. However, if an effect does not respond to replicators' call, the weight of evidence shifts against it or, worse, its existence is cast in doubt and void, as has recently been done with regard to ego depletion, social priming, bystander effect, actor-observer asymmetry in attributions, loss aversion, delay of gratification, and other phenomena (Malle, 2006; Doyen et al., 2012; Carter et al., 2015; Gerber et al., 2016).

Whether a phenomenon has truly revealed itself is decided by statistical means, typically *p*-value. It can work in physics (Meehl, 1967) where, for example, multiple experiments in Switzerland revealed the odds of one in 3.5 million in favor of the existence of the Higgs boson particle (or that the result would occur if the null hypothesis were true). In psychological studies of human behavior, however, effects of such magnitude and precision do not exist, as the same experimental treatment can produce a *p*-value of 0.001 today but 0.75 tomorrow (Cumming, 2014).

Besides the statistical problem, the yes-or-no determination is logically untenable because the absence of evidence derived from a replication is taken to indicate that a phenomenon does not exist (Carter et al., 2015; LeBel et al., 2017). The absence of evidence can result from many constraining factors, methodological and measurement factors on one hand and time- and context-related determinants on the other, and cannot therefore be taken as evidence for the absence of a truth or phenomenon (Trafimow, 2003). Moreover, because empirical support is always provisional and propositional, and therefore preliminary, conditional, and relative, the categorical determination of scientific truth is not possible. Accumulating evidence from phenomenon replications, however, provides a better understanding of the phenomenon and its temporary truth-value.

The preponderance of evidence for a phenomenon only provides a more probable or justified explanation than other explanations at the present time (Kuhn, 1962; Meehl, 1990), but offers no final truth (McFall, 1996). A current theory or explanation has not yet been shown false or has not been disproven or "falsified" (Popper, 1959), nor have alternative hypotheses been accepted by "strong inference" (Platt, 1964); for a good example of strong inference ruling out alternative explanations in empirical research, see Oppezzo and Schwartz (2014).

MEASUREMENT AND REPLICATION

In science, the nature and acceptance of truth is importantly shaped by measurement. This means that replications are essentially about measurement invariance and reliability of previous findings, especially if studies aim at exact replications. In principle, if a phenomenon is successfully replicated, it is viewed as a *reliable* and bona fide effect, and a scientific truth, albeit temporary, is therefore established. However, measurement invariance is not just a matter of reliability, but validity as well (Hussey and Hughes, 2020). That is, does a successful replication automatically capture the underlying mechanism of the investigated phenomenon? Was the underlying or latent construct measured validly?

While reliability is important, a more critical issue is validity. An original finding can be replicable but nevertheless lead to invalid conclusions, because replicability is not the same as validity (Hussey and Hughes, 2020). If we can measure something reliably but it is off the target, such a replicated finding means little for understanding the underlying phenomenon. In other words, what does a successful or unsuccessful replication with an invalid measurement or test mean for scientific truth? If, for example, behavioral tasks are used as dependent measures to capture the mechanism underlying self-regulation, such a research strategy is problematic because behavioral tasks have been shown to possess a relatively low test-retest reliability (Enkavi et al., 2019). As reliability and validity are interrelated, lower reliability leads to lower validity, meaning that behavioral measures are less accurate and valid for measuring the underlying mechanism for self-regulation (Enkavi et al., 2019). Low reliability, and thus low validity, of behavioral measures increase replication failures.

The problem of validity is also a matter of poorly defined theoretical constructs, resulting in a problem of overlapping constructs and measurement variance (Hanfstingl, 2019). For example, there is considerable theoretical overlapping between such constructs as "grit," resilience, self-control, mental toughness, and "self-as-doer." In a similar vein, a single term or construct can have different meanings. Smiling as a response to a stimulus may be replicated reliably, but it often has different meanings in different contexts.

Given that psychological scientists seek to provide robust explanations for various phenomena through latent constructs, measurement validity becomes a critical issue (Hanfstingl, 2019). Thus, replicable findings are useful only if they reflect differences in latent variables, not just how reliably participants interpret items in the questionnaire (Hussey and Hughes, 2020). From the validity standpoint, then, experimental participants' performance should be driven by the underlying construct, such as physical fitness being the construct determining the treadmill test performance (Secrest, 1984). A resultant number has a specific meaning that represents the underlying construct, reflecting its construct validity (Cronbach and Meehl, 1955). In this common quantitative approach, validity is conceptualized as a matter of degree, not as qualitative concept of "yes" or "no" (Cronbach and Meehl, 1955; Messick, 1989).

Although it is often thought that reliability sets the upper bound for validity (Lord and Novick, 1968), technically speaking, this is inaccurate because the maximum validity of a test is the square root of the reliability (Secrest, 1984). Nevertheless, increasing measurements' reliability in original and replication studies is necessary; unfortunately, it is not uncommon to find reported reliability coefficients to be less than the recommended standard of 0.70 (Nunnally, 1978). For example, in task-fMRI studies, measures' average test-retest reliability (0.397) is very poor, making them unsuitable for brain-behavior mapping for both research and clinical purposes (Elliott et al., 2020). In sum, successful or unsuccessful replications based on highly reliable and highly valid measurements will appreciably add to the existing knowledge. If, on the other hand, the original study and follow-up replications were based on relatively unreliable and invalid measures, such studies would have little, or no, value in the establishment of scientific knowledge and truth.

Even in physics the present scientific truth is subject to revisions due to measurement problems and use of different methods. For example, two groups of scientists have arrived at vastly different numerical values for the rate of the expansion of the universe, with one indicating that the universe is receding about 9% faster than the other (Panek, 2020). A great puzzle among astrophysicists is whether the discrepancy is due to a systematic error in one of the two measurements or whether a "new physics" (e.g., dark energy changing over time) altogether is needed to explain the "inflation" of the universe. In a similar vein, the beginning of the universe itself has been questioned by some physicists, who have suggested that Big Bang may not be taken as an unchangeable conclusion and truth; accordingly, Big Bang could actually be "Big Bounce" in a constellation of infinite universes ("multiverses").

By the same token, there is no denying that certain laws of the nature are permanent and fixed entities. For example, light always travels at the constant speed and faster than anything else in a vacuum, even if it slows down to 75% of the vacuum speed in water. There are, however, no such permanent laws in psychology because psychological phenomena are not fixed and unchangeable particles. Yet stability (e.g., speed-accuracy tradeoff) exists in human behavior. As stable patterns are tendencies, not laws, in human affect, cognition, and behavior, they become less stable under certain conditions. This is a challenge for empirical science in psychology in general and for making generalizations about human behavior in particular.

The above suggests a major difference between physics and psychology. While the laws of nature are fixed and stable entities, though subject to revisions, psychological phenomena are a mixture of both variability and stability. In illustrative terms, when light travels at its constant speed, it does not have an ability to slow down at Jupiter to admire the scenery, whereas the human mind does it in various contexts, processing information differently as a function of internal and external conditions. Because humans are capable of changing and evolving, their feelings, cognitions, and behaviors can fluctuate substantially, but at the same time, there is a considerable degree of stability to them (Hudson et al., 2017). In fact, at the fundamental level, attention and visual search are biased toward temporal regularities (Zhao et al., 2013). The upshot is that the establishment of scientific truth is as hard, if not harder, in psychological as in other sciences. There are no invariant particles to be discovered beyond a shadow of doubt in psychology, but instead, variable, temporary, context-dependent, and subtle phenomena (Iso-Ahola, 2017). As a result, replications can only provide experimental feedback for hypotheses and theories, but not declarations that certain phenomena are not real.

LAB TRUTH VERSUS REAL-LIFE TRUTH

The replicability problem is evident when considering that the nature of psychological phenomena is intertwined with their methodological/measurement demonstrations. Accordingly, self-control prevents people from using profanities in public (equivalent to lab situations), but not when they interact with trusted friends in private settings (real-life situations). Field experiments, while important for external validity, tend to compromise internal validity (Shadish et al., 2002) and make the findings of such studies difficult to replicate. Because intervening variables are hard to experimentally control in real-life situations, the measured influence of these variables will not be identical in original and replication studies. The tradeoff between internal and external validity is nothing new but should not be ignored, because it suggests a marked difference between lab truth and real-life truth.

In theory, phenomena exist in lab settings but not in real life, and vice versa. This yields four scenarios according to where replications have been conducted and whether replications have succeeded or failed: (1) successful replication in lab, (2) successful replication in real life, (3) unsuccessful replication in lab, and (4) unsuccessful replication in real life. Following the first case, a question arises: Is this lab phenomenon a bona fide effect in real life as well or just limited to lab situations as a methodological artifact? Although reaction time, and its determinants, can reliably be replicated in labs, it may be less replicable in real-life situations (e.g., a slowed reaction time in traffic due to fear). Similarly, causal attributions can be elicited in labs by asking participants to causally attribute their performance outcomes, but in real life, people rarely engage such conscious thoughts.

Regarding the second situation, a question is whether these successful real-life replications make the phenomenon more credible, as in the case of replicated social priming effects in a variety of real-life situations with real-life variables (Bargh, 2014), even if some lab experiments have failed to replicate them. It could be argued that the greater the number of real-life situations in which the phenomenon has been replicated, the greater the likelihood that it is a bona fide effect, especially if the empirical demonstrations are consistent with lab replications. In general, empirical demonstrations and replications in both lab and real-life settings provide the strongest evidential support for phenomena, whereas a lack of both lab (situation 3) and field confirmation (situation 4) provides the least support.

In the third case, the phenomenon does not exist as a lab effect but raises questions: Does the absence of lab evidence rule out methodological artifacts as the reason for replication

failures, and do these failed lab replications necessarily rule out the phenomenon's viability in real life? Ego depletion may a weak phenomenon when participants perform behavioral lab tasks, but it could be a strong effect in everyday life contexts (Iso-Ahola, 2017). In the fourth situation, evidence is lacking in both labs and field settings and therefore suggests an end to scientific inquiry on the phenomenon. However, it is possible that methodological problems and artifacts have failed to reveal the phenomenon (e.g., cognitive dissonance) in real-life situations.

The four scenarios suggest a complex and non-categorical role for replication in psychological science. Scientific truth is stronger when it is based on corroborated evidence obtained from lab and real-life situations. However, some have described psychology as the science of self-reports and finger movements and have asked, whatever happened to actual behaviors (Baumeister et al., 2007)? It is therefore not surprising that cognitive neuroscientists, for example, have recently called on researchers to investigate the relationship between the brain and human behavior in the context of everyday experiences (Shamay-Tsoory and Mendelsohn, 2019). Such "ecologically valid" studies are to be based on "representative designs" in which both individuals and stimulus situations are representatively sampled for better understanding and generalizability (Brunswik, 1955). Real-life human behavior is always dynamic, interactive and complex, and therefore poses major challenges for controlled experimentation. Nevertheless, real-life replications are just as important as lab replications for providing cumulative and convergent evidence for various phenomena.

STABILITY AND VARIABILITY

That a need for the balance between stability and variability is deeply embedded in human nature and manifested in behavior is well documented. For example, college students spend about half of their time doing the same things day after day, but the flip side is that they spend the other half doing different things (Wood et al., 2002). Thus, they seek both stability and variability, consistency and novelty, in their lives. According to Berlyne (1960), this tendency is driven by the need for optimal arousal. The need for variety (and stimulation) is so great that experimental participants find sensory deprivation conditions intolerable, as illustrated by their request to hear a recording of an old stock market report over and over (Murray, 1964). This raises a question for experimental research: Is participants' need for variability and stability met to the same degree from one study to another? Are participants under- or overstimulated to the same extent?

Overall, both emotions and cognitions have stability and variability to them, be they implicit or explicit attitudes (Charlesworth and Banaji, 2019), implicit biases (Vuletich and Payne, 2019), or day-to-day affect (Hudson et al., 2017). Recent evidence further indicates that affect variability, not just the mean stability of positive affect, is associated with health-related physiological outcomes (e.g., antibody response to vaccination) (Pressman and Cross, 2018). In short, it is not surprising why replications of certain psychological phenomena have failed; they have in part failed because it has been assumed that participants are affectively and cognitively stable from experiment to experiment and their need for variability and novelty does not influence their experimental performance. Without measuring these constructs and their proxy variables, it is not possible to determine the extent to which the basic need for stability and variability contributes to differences between experimental findings.

The inherent variability of psychological phenomena can further be influenced by contextual and time-related factors (Ramscar et al., 2015; Bavel et al., 2016; Ramscar, 2016). Every experimental situation is a different social context and can interactively with time affect participants' behavior. Psychological phenomena do not occur in frozen time and context but vary substantially as a function of these determinants. Human behavior is dynamic, continuous, and interactive, even if it is studied in frozen contexts and at frozen times in research labs. Furthermore, even if people tend to think and act non-consciously most of the time (Bargh, 1997; Dijksterhuis and Aarts, 2010), they have cognitive capacity to change their feelings and thoughts in any place at any time (Baumeister et al., 2011; Dehaene, 2014). Experimental settings and instructions could easily give rise to such changes (Belletier et al., 2019), which suggests that it is imperative to follow verbatim experimental instructions and scripts in replication studies (and include them in publications). Phenomenon replications are therefore needed to examine the effects of deviations from and variations in experimental details on replication results.

In efforts to create equally homogeneous external conditions, original studies and their replications seek to constrain lab conditions to the point where experimental participants are stripped of their very psychological being and behavior. Yet, in everyday life, psychologically functioning individuals have freedom to feel, think, and act as they desire in specific situations at specific times, thereby creating complex interactive relationships. These internal conditions, however, cannot be made entirely or even sufficiently homogeneous from one lab to another and yet, it is these internal conditions in conjunction with external factors that determine variability in participants' responses. They (e.g., conscious thoughts) can be suppressed by strict and artificial methodological requirements, but it is unclear whether such suppression is the same between the studies.

Thus, it is not known if the replication participants have had identical conscious feelings and thoughts of excitement or boredom, for example, to say nothing about non-conscious feelings and thoughts. In the reported replications, no attempt has been made to create internal conditions homogeneous relative to the original conditions. There is not a single replication study reported that would have shown participants' feelings and cognitions—both conscious and non-conscious—to be precisely the same with those of the original participants. Although people generally are inclined to rely on non-conscious processing of their thoughts, emotions, and behaviors (Kahneman, 2011; Bargh, 2017), they can deviate from this tendency at any time (Baumeister et al., 2011), and experimental instructions can produce different degrees and ratios of conscious vs. non-conscious processing from one study to another.

These degrees and ratios, however, can be estimated and should be taken into account to separate real effects from extraneous influences. It is proposed, on the basis of a vast literature on conscious vs. non-conscious processing (Kahneman, 2011; Bargh, 2017), that the greater the degree of conscious processing required for performing an experimental task, the less replicable are the findings because of increased within-person and between-person variance; and in reverse, the greater the role of non-conscious processing, the more replicable are the results.

It should be noted that some experimental tasks (e.g., a right-handed person doing something with his/her left hand) require greater cognitive awareness and resources than others, and that most tasks become increasingly non-consciously processed with repeats (Bargh, 2017), thereby diminishing cognitive fatigue (Hockey, 2011). This suggests that a replication may have failed because the experimental task, instructions, and manipulations led participants to engage in different degrees of conscious vs. non-conscious processing and use of cognitive resources or because the phenomenon is inherently associated with one type of processing vs. the other. For example, do people get more ego-depleted when doing consciousness-demanding vs. non-consciousness-demanding tasks?

The complexity of the human mind means that psychological phenomena are essentially interaction effects (Mischel, 1973), products of time and social context, thereby increasing the likelihood of failed replications. Both the beauty and challenge of the human mind for researchers is that people's thinking varies from simple to complex, from what might be called simple "main-effect thinking" to complex "interaction-effect thinking" (Iso-Ahola, 2017). In their thinking (conscious or non-conscious), people can be simple at one time and complex at another, or both at the same time. They can be either rational or irrational or both in the same situation (Tversky and Kahneman, 1971; Lerner et al., 2004; Lieder et al., 2018).

How does a replication researcher know the mode of thinking in which his/her participants are vis-à-vis the participants of the original study? In short, it cannot be assumed that experimental instructions will make every participant think, both consciously and non-consciously, in the same way. This emphasizes the necessity of reporting precise experimental instructions in publications. Such information would allow researchers to identify within- and between-subjects factors that cause individuals to react differently to instructions.

The complexity of psychological phenomena means that many variables will interactively influence outcome variables, which poses major problems for replications. As a whole, other factors being equal, the more complex and interactive the effects, the less replicable and reproducible they are. Meehl (1990) famously concluded that "everything is correlated with everything, more or less." An important implication is that the manipulation of a focal independent variable influences other causal independent variables even in randomized experiments, and these other independent variables either reduce or increase the effect of the focal variable. Such interaction effects can invoke divergent thinking in participants and thus lead to more variability in behavior from time to time, from situation to situation, which in turn reduces the probability of successful replications of the original findings, as well as predictability of behavior.

TEMPORARY TRUTH

Whether pursued through inductive or deductive reasoning and methods, scientific knowledge is subject to continuous revisions, resulting in temporary and preliminary truth. Since investigated effects are time-bound, there are only "temporary" winners in knowledge and no final truths (McFall, 1996). A good example of the temporary nature of psychological knowledge can be seen in the virtuous cycle of daily motivation. As people's senses of strivings are dynamic, they change from day to day so that today's need satisfaction at work can enhance the next morning's strivings and work behaviors (Foulk et al., 2019). How, then, can replication studies be expected to reproduce the original participants' feelings, thoughts, and acts when they change from one working day to another? Can it be assumed that these factors vary, similarly, from study to study, especially when participants are typically not randomly selected? What is the underlying pattern here to be replicated? Is it that people are dynamic, changing individuals, or is it just motivation?

If replications fail to reproduce the original findings, it could easily be because of the dynamic cycle of daily motivation and the measurements taken at different times of the day. Related to this point, the effect of coffee on cognitive performance may depend on the time of drinking (morning vs. afternoon) (Ryan et al., 2002), and the effects of verbal cues on voting behavior are contingent on whether measurements are taken 1–2 days vs. 3 days prior to actual voting (Bryan et al., 2011; Gerber et al., 2016). All of this again underscores that humans are complex and variable beings, and this complexity and variability therefore needs to be accounted for theoretically with parsimonious models (Saylors and Trafimow, 2020), and empirically with high-powered studies in labs and real life.

Another thing to be noted is that psychological effects occur both in short and long periods of time. In the latter, the effects often become reciprocal, as demonstrated in the reciprocity of prosocial behavior and positive affect in daily life (Snipper et al., 2018), reciprocity of self-control depletion and self-control practice (Baumeister, 2015), reciprocity of athletic performance and the perceived size of softball (Witt et al., 2012), and reciprocity of self-concept and athletic performance in a specific performance situation (time) and cumulatively over time (Marsh and Perry, 2005). Thus, a replication experiment may fail to reproduce an original finding because it is a frozen snapshot of one-way effect in time when the effect actually fluctuates with time due to a continuous reciprocal relationship between the variables. In a game situation, for example, performance continuously affects the perceived size of the ball, which in turn continuously affects performance (Witt et al., 2012). If a field experiment is replicating the former relationship but ignores the latter relationship, the replication is likely to fail. For this reason alone, it is not logically possible to declare that a phenomenon is permanent or that it does not exist.

It should also be noted that although one-time lab exposure to a stimulus (e.g., meditation training) can produce an observable effect, long-term exposures are more likely to result in more replicable effects because they solidify and reinforce single-exposure or training effects (Kok et al., 2013). In medical science, for example, the establishment of the efficacy of a treatment often requires long-term exposures to drugs. Similarly, environmental toxins may take years to produce harmful effects (e.g., smoking and lung cancer). Even if the findings of the original study were obtained from a short-term exposure to a specific toxin, the effect is more likely to be replicated following repeated exposures to it. However, there is no guarantee that a toxin as a single causal factor will lead to successful replications. As noted earlier, the manipulation of a focal independent variable can influence other causal independent variables.

In general, the same principle applies to replication of psychological effects. Because human behaviors are products of multiple time-bound and context-bound factors, and are therefore products of interaction effects, psychological phenomena can only increase the probability of the occurrence of certain responses and actions. Thus, a psychological effect does not itself cause changes in a behavior but interactively contributes to them. For example, it is known that social priming produces differing motivational effects rather than a single default effect (Loersch and Payne, 2011). The absence of such one-time generalizable effects inevitably reduces the replicability of original findings. The situation calls for better and more accurate theories and models.

Finally, time could play a different role depending on the degree of proximity to major social and religious events and periods. If, for example, original participants' favorite sport team just lost a championship game, they would likely be in a different mood than replication participants who recently had more pleasant experiences. Is it known that multi-lab and multi-culture replication participants match the original participants on such accounts? In addition to developing better theoretical models to account for this potential variability, methodological and statistical solutions involve large sample sizes, mixed-effect models, and treatment of lab or culture as a factor in replications.

CONTEXT-BOUND TRUTH

Psychological truth is not only time-bound but also context-bound. In general, unlike particles (e.g., photons carrying light), humans do not function and operate in vacuum or optimal experimental conditions but instead in social environments, influencing them and being influenced by them. Contexts may be similar but never identical. Although physical contexts can be reproduced, identical social situations cannot be re-created. For example, spectators go to the same stadium to watch a football game on consecutive Saturdays, but every game is a different social context for many reasons, even though the football game itself is a constant from Saturday to Saturday. Needless to say, no replication can duplicate a previous game's psychological context, which means that psychological truth is relative to contexts in which a behavior occurs. Since contexts change with time, psychological truth is variable by its nature. How to quantify and predict this variability becomes a key theoretical and empirical issue.

Thus, the same degree of anxiety can cause "choking" (deteriorated performance) in one situation but not in another. Notwithstanding methodological problems, replication failures likely reflect the investigated phenomenon's inherent variability due to its sensitivity to situational influences (Bavel et al., 2016). This contextual sensitivity also means that instead of making generalized claims, original findings have to be viewed with empirical limitations and uncertainty in mind until the conditions reliably producing certain effects are well understood. What factors make various phenomena more or less sensitive to environmental influences, and thus more variable than stable, remains to be examined both theoretically and empirically.

As time and context are inextricably intertwined, human behavior varies as a function of the interaction of the two, with one affecting another. Contexts vary with time as time creates different contexts, and contexts (e.g., boring vs. exciting lectures) create different experiences of time. Christmas Day is a time that affords a different context and experience than New Year's Day. With time and context, internal and external conditions change and are therefore different in a replication study when compared to the original.

Time-bound and context-bound effects mean that psychological truth is also person-bound. In other words, psychological truth has intrapersonal boundaries. Intrapersonal differences suggest that the same person is not always similarly affected by time- and context-related variables. There is variability and stability within a single individual's responses, while interpersonal differences indicate that people differ from one another in the degree to which time and context influence their affect, cognition, and behavior. Psychological studies are predominantly conducted to determine between-subject variance while ignoring within-subject variance. In this approach, intrapersonal (in)consistency gets statistically buried in interpersonal (in)consistency, with weak between-subject averages potentially hiding strong and meaningful intrapersonal effects. However, the problem can in part be remedied by using mixed-effects models to account for between-subjects variance, as well as within-subjects variance by treating stimuli as random effects.

To better understand psychological effects, variability and stability need be investigated and replicated at both intrapersonal and interpersonal levels. In medical research, between-subject determinations (experimental vs. control group) of a new drug's efficacy may indicate lifespan extension, on average, by only a few months, but individually the effect could be several years for some. Although psychological experiments do not directly deal with life-and-death effects, they need to pay more attention to intrapersonal stability and variability of various effects.

SUBTLE TRUTH

Scientific truth can be subtle and elusive. The first discovery of gravitational waves (2015), as predicted by Einstein's theory, was

more or less accidental and could easily have been missed or misinterpreted. Similarly, a particle presumed to constitute dark matter continues to elude physicists, a situation not uncommon to psychologists. Nevertheless, "a good deal of good physics has been done without high quantitative precision" (Chang, 1995, p. 169). There is no reason why the same could not be said about psychology, even though a notable difference between physics and psychology is that psychological phenomena are not particles that prevail across time and situations. Instead, they vary in time and space and therefore become elusive and subtle, and open for misinterpretations.

Consider, for example, the human face. It can give subtle signs of different emotions as responses to certain stimuli, but those signs can be different in different situations at different times even when exposed to the same stimulus. In laboratories, participants have been able to make reliable judgments from government officials' faces regarding their likelihood of having been convicted of corruption (Lin et al., 2018). Although such "main effects" may be replicated, interaction effects (i.e., if target persons' talk and non-verbal behavior are included in the analyses) may not.

Or, consider sarcastic comments. In a subtle way, they can communicate different messages, which can easily be misinterpreted by others. These kinds of subtle effects can lead to the problem of "inferential reproducibility" (Goodman et al., 2016), meaning that different scientists draw different conclusions from the same results. Such inferential errors abound in replication studies of various phenomena, from the effects of subtle linguistic cues (Gerber et al., 2016) to the effects of ability vs. practice on expert performance (Ackerman, 2014; MacNamara et al., 2014). These errors have been particularly common in so-called failures to replicate social priming effects. But as it has been pointed out, associations between priming words and the world are not stable over time because of learning and experiential effects (Ramscar, 2016). It is therefore not surprising that many exact replications of social priming have failed and will fail statistically (Ramscar et al., 2015).

Subtle effects do not necessarily mean that they are not noticeable and important effects. For example, the tone of one's voice can in a subtle way communicate strong positive or negative feelings to others, yet statistical data analyses could indicate weak effects. Small effect sizes and low multiple correlations can be indicative of important effects, and an unsubstantial manipulation of the independent variable can produce non-trivial effects (Prentice and Miller, 1992). Furthermore, subtle variations in experimental settings from one study to another (e.g., the experimenter's behavior) "may cause dramatic changes in performance" (Belletier et al., 2019), resulting in conflicting findings and failed replications. Similarly, experimenter beliefs rather than the primed condition can alter participants' behavior (Gilder and Heerey, 2018). It may therefore be proposed, other things being equal, that the subtler the effects, the greater the likelihood of replication failures; however, such failures cannot necessarily deny the underlying effect. In short, it is clear that the subtlety of psychological effects on one hand and their sensitivity to social contexts on the other pose major problems for replications.

To be sure, some phenomena are manifested in strong and thus potentially more replicable effects, as has been shown by research dealing with the influence of the mere sight of individuals with a gun on subsequent aggressive thoughts (Bushman, 2018). By analogy, vigorous exercise produces stronger physiological effects than moderate and mild exercise. Rarely, however, do people in everyday life function under extreme psychological conditions, nor are the psychological effects likely to be linear, as in the case of exercise intensity. Moreover, the magnitude of an effect is only one criterion by which a phenomenon's viability can be assessed, and in fact, it may be a poor or inappropriate criterion in many situations. Phenomenon-focused replications would look at an effect not just in terms of its magnitude but also by its frequency, duration, and intensity in labs and real life, as well as whether the phenomenon (e.g., stereotypes) is consciously or non-consciously experienced (Iso-Ahola, 2017). It is known that most psychological phenomena become non-consciously experienced emotions, cognitions, or acts with time (Bargh, 1997, 2017).

Subtle effects may not be quantitatively strong and yet be qualitatively and theoretically meaningful. It has been shown that small effect sizes can potentially be important for both theoretical and practical or utility purposes (Abelson, 1985), as in the case of the effects of mild exercise on cardiovascular health. However, given that psychological studies are statistically underpowered (Rossi, 2013) mainly because of small sample sizes, they are not capable of detecting small effects; large samples are needed to detect small (and subtle) effects. It should be noted that although increasing sample size increases informativeness and power, "high power does not necessarily imply an informative or high-quality experiment" (Cumming, 2014), nor can the average power serve as a "replicability estimate" (McShane et al., 2020).

Although strengthening experimental manipulations increases effect sizes, it can undermine effects' subtlety. The other side is that because of ethical and moral reasons, some psychological variables cannot experimentally be manipulated to the level at which they occur in real life. As a result, weak lab manipulations are likely to produce weak effects that do not replicate. To counter these problems, phenomena need to be investigated in a diversity of experimental and real-life tasks and situations with different types and degrees of manipulations. In the long run, such constructive replications will yield convergent evidence rather than yes-no determinations of phenomena's existence.

It would be a mistake to demand that subtle psychological phenomena produce linearly and quantitatively increasing effects before they are deemed real. Such statistical demands would ignore the complex, interactive, non-linear, and subtle nature of psychological phenomena and would reduce them to statistical phenomena. Stringent statistical demands create an illusion of objectivity but do not eliminate subjective questions about "inconsequentially small" but important effects or their "practical meaning," as has recently been debated regarding the long-term effects of delay of gratification on behavioral outcomes.

Statistical decisions are not error-free either. It has been suggested that some replication failures may be due to

applications of the wrong type of Type I error testing procedure: Neyman-Pearson vs. Fisherian Type I error rate (Rubin, 2019). New statistics are no panacea either. Meta-analyses (effect sizes), for example, have led to contradictory findings regarding various phenomena (Hagger et al., 2016; De Vrieze, 2018), although they can be useful in suggesting moderating or interactive effects. In the end, however, statistics cannot yield automatized yes-no decisions on substantive hypotheses' believability because "there will always remain an important role for expert judgment" (Trafimow, 2019).

In sum, the subtlety of psychological phenomena and their sensitivity to social influences inherently cause variations in empirical observations and pose challenges for reproducibility. There is an important difference between a phenomenon's subtlety and its observed statistical strength. Feedback, for example, can be given in a subtle or concrete way, but it is only in the latter situation that feedback is likely to show statistical strength. A danger is that a phenomenon's subtlety will be lost if it has to be manipulated to an artificial statistical strength, turning naturally occurring subtle effects into statistically strong but theoretically questionable effects. A long-term danger is that psychological phenomena will increasingly become defined as statistical phenomena.

CAUSALITY AND TRUTH

The elusive and complex nature of psychological phenomena is reflected in their underlying causes. Let us consider the much-investigated "expert performance" as a phenomenon. Are "talent" (Ackerman, 2014) and "deliberate practice" (Ericsson et al., 1993) necessary or sufficient for exceptional performance? Evidence indicates that each is necessary but not sufficient for producing expert performers in any area of human performance. Most psychological phenomena, however, are not based on yes-or-no necessary causes (e.g., the female gender being necessary for pregnancy), but when they are, it is likely that necessary causes are more replicable than sufficient causes.

Although talent is necessary for expert performance, its contribution varies from one performance domain to another, and thus potentially from one replication to another, depending on how performance is defined and measured (Ackerman, 2014). Even if replications would show talent's contribution to vary in percentage and be weaker than that of deliberate practice, such replications could not deny the basic fact that talent is necessary for expert performance. Typically, however, psychology experiments seek to establish sufficient causation, as follows:

A. If **X** (self-control resources exhausted during a working day) occurs and **Y** (ego depletion effect) is observed, then the **X**-**Y** causation is true.
B. If **X** does not occur and no **Y** is observed, then the **X**-**Y** causation is indeterminable.
C. If **X** occurs but no **Y** is observed, then the **X**-**Y** causation is untrue.
D. If **X** does not occur but **Y** is observed, **Y** is true but independent of **X**; no **X**-**Y** causation.

In the above scenarios, A and B, when taken together, seek to establish sufficient causation, in that when **X** is present it produces **Y**, but when it is absent **Y** does not occur. This is the usual treatment-control group situation in experimental research, but it is only one-time demonstration of sufficient causation. Further, an experimental confirmation of the effect of **X** on **Y** would not establish the necessary causation, according to which **X** *must* be present for the effect to follow. It is possible that **X** is necessary for **Y** to occur, but a researcher does not know it from one experiment; a series of experiments would have to substantiate it. And, if the D scenario is encountered, it would deal the death blow to the idea of necessary causation. Using the above example, the ego depletion effect (self-control failure) was observed not because of **X** but because of other factors, such as physical or mental tiredness or habit-supporting cues (e.g., a remote control prompting TV watching). These other factors can empirically be shown to be sufficient causes for **Y**'s existence.

Finally, the C scenario would represent a failed replication of the effect of **X**, suggesting a lack of sufficient causation that whenever **X** is observed it will lead to an effect. Given that psychological effects fluctuate with time and context as a function of internal and external conditions, the refutation of sufficient causation by replication becomes untenable in logic and reality. Furthermore, if there is not a single cause whose presence will always lead to the effect, the entire premise of sufficient causation is in question. In general, the difficulty of empirical verification (replication) of causation stems from the fact that there are no single causes that are both absolutely necessary *and* absolutely sufficient (for a more detailed discussion of probabilistic causation regarding intervention effects and "the probability of sufficiency", see Pearl and Mackenzie, 2019).

To further elucidate the problem of demonstrating necessary and sufficient causation in psychological research, let us consider research on delay of gratification. Is the ability to delay gratification in childhood necessary and/or sufficient for better behavioral outcomes 10 years later? Clearly, it is not necessary because people can succeed without delay of gratification, nor is it sufficient because delay of gratification does not guarantee the effect (i.e., it does not always lead to successful performance). Several other factors (e.g., family resources and income) can result in successful performance many years later. Nevertheless, delay of gratification can increase the probability of success for certain groups of individuals and under certain conditions (Mischel et al., 1988, 1989; Casey et al., 2011), especially if it means enhanced perseverance and associated deliberate practice.

Any researcher knows that it would be imprudent to causally ascribe today's performance to an experimentally exhibited behavior 10 years earlier, because there are numerous intervening variables in the span of 10 years that could easily affect today's outcomes. In fact, it would be remarkable if the correlations between one variable measured 10 years ago and certain indicators of success today would not be low. These low correlations, however, do not constitute replication failures as such and do not therefore give a license for researchers to throw out the baby with the proverbial bath water (Watts et al., 2018). Low correlations and small effects can be important and meaningful in shedding light on underlying phenomena

(Prentice and Miller, 1992). In other words, does anybody seriously think that delay of gratification, or perseverance more generally, is not important for successful human performance? If he/she does, he/she then needs to explain why more than 10,000 h of "deliberate practice" is required for becoming an exceptional performer (Ericsson et al., 1993).

Research on delay of gratification illustrates the difficulty of being able to draw hard empirical conclusions about psychological phenomena and their truth-value. For one thing, seldom or few psychological phenomena are stand-alone effects. Rather, they exert their influence through other factors, which then makes the effects more variable and less replicable, but not less important. For another, because all psychological effects are time-bound and context-bound, their causes are not exclusively either necessary or sufficient as such. For example, although anxiety is not necessary for "choking" (impaired performance), it can be sufficient under many conditions (i.e., in certain context and time). But establishment of sufficient causation between **X** and **Y** with longitudinal data (e.g., research on delay of gratification) poses major methodological and statistical challenges for replications of the original findings.

Rather than being surprised or dismayed by so-called replication failures due to the complexity of psychological effects, they should be embraced not as the final arbiters of scientific truth (no yes-or-no determinations) but merely as additional feedback in the ongoing theoretical and empirical enquiry of multifaceted psychological phenomena. Replication failures can be useful for revealing phenomena's boundary conditions and informing researchers on the interaction effects of context and time, and individual differences.

More generally, replication failures can aid in developing theories that last for the longest amounts of time over the greatest ranges. Parsimonious theories are such theories (e.g., Newton's theory) and can be achieved when unimportant or non-influencing variables, overlapping constructs, and exogenous causes are removed, when auxiliary assumptions linked to theory (Earp and Trafimow, 2015) and specified theoretical effects (Witte and Zender, 2018) are tested and replicated, as well as when the relationships are properly theorized instead of adding more variables to an ever-growing list of moderators and mediators. Evidence indicates that "complex models are much less likely to be true than simple models" (Saylors and Trafimow, 2020).

REPLICATION AND PSYCHOLOGICAL PHENOMENA

Understanding the role of replication in the scientific process is rooted in the assumption that the identical method has to produce identical results. As argued, this assumption is false because methods themselves are psychometrically limited and because they are not identical from one study to another. In addition, as psychological phenomena exist in different forms, degrees, and contexts, they necessitate employment of the same method in different situations and different methods in the same situation. Failures to conduct replications under this principle have led to unjustified conclusions and assertions, as well as denials of many classic phenomena. Sweeping generalizations have been propagated from the attempted yes-no replications, as was the case with the BBC television prison study claiming to replicate Zimbardo's Stanford Prison Experiment. The replication bore little resemblance to the original in the employed experimental protocol, not to mention any consideration of other studies on social power. Focusing on Zimbardo's method, questionable or not, is losing sight of the main thing, the phenomenon itself: how social power affects human behavior.

Regardless of disagreements on the methodological procedures, Zimbardo's and Milgram's experiments undeniably showed the power of social influence on individual emotions, thoughts, and behaviors in the reported experimental situations. Even when considering the special lab conditions created in these experiments, no replication can deny the fact that Milgram was able to create an experimental situation in which social power made participants ostensibly hurt other humans. In other words, the obedience effect was observed in that specific situation at that specific time, and it therefore existed then and exists today, even if in a different form and degree.

These original findings do not mean that social situations overpower every individual at all times, as shown by Milgram (1963) himself. While 65% of participants in his experiment gave the highest level of electric shock to the confederate, 35% did not. This demonstrates the essence of psychological phenomena and their empirical verification: studies may discover patterns and replications may confirm them, but patterns, by definition, do not cut across time, situations, and persons. They vary not just between individuals but also within individuals. The problem with patterns is that they are quantitative and statistical patterns from which the operation of psychological effects is inferred. Increased statistical power and larger sample sizes certainly increase the credibility of patterns, but they do not eliminate the "inferential reproducibility" problem (Goodman et al., 2016) that different researchers tend to draw different conclusions from the same data.

There are numerous examples in the literature of methodological liberties replication researchers have taken in attempts to disprove original findings (Doyen et al., 2012; Gerber et al., 2016). The "replicator degrees of freedom" have been shown to lead to unwarranted claims of replication failures (Ramscar, 2016; Bryan et al., 2019). Thus, it cannot be assumed that so-called independent replications are unbiased. In fact, it has been suggested that there is an incentive to find the null result refuting the original finding (Bryan et al., 2019). Refutations can readily be obtained by liberal uses of replicator degrees of freedom [e.g., prior selection of experimental designs and analyses (Bryan et al., 2019; Sherman and Rivers, in press)]. Furthermore, people's general sensitivity to social influences (Bavel et al., 2016) makes it relatively easy to conduct, unwittingly or not, replication experiments to produce refuting evidence. One way to safeguard against methodological biases would be for replication studies to provide well-developed theoretical rationales that would specify beforehand under what conditions certain effects are likely or unlikely to be found and replicated. Such theoretically based

replications would make greater contributions than mere methodological replications.

Regarding the classical studies of psychology, it should not be forgotten that they were original discoveries or demonstrations of various phenomena, which naturally is more important than any replications of them. If Festinger (1957) did not discover cognitive dissonance, there would be no cognitive dissonance to be replicated. Thus, initial theorizing is critical for the advancement of science, and innovation should be promoted and not suffocated by the overemphasis on replication, as Lancet's editorial (2017) expressed it: "Prescriptive regulation of scientific thought and processes that stifle creativity under a guise of enforcing reliability could ultimately impede discovery and advancement." This argues for a balance between theoretical innovation and empirical research, including phenomenon and constructive replications investigating potential confounds and testing competing explanations and specified theoretical effects for the generalizability of the original findings (Witte and Zender, 2018).

PSYCHOLOGY AS SCIENCE

Although so-called replication failures have raised questions about psychology as a serious science, good news is that in the long run, the value of scientific psychology cannot be diminished by any of it, for one thing, because psychology is the only science that can answer some important questions about human life. For example, why is it that most people do not exercise regularly even though physiological evidence has compellingly demonstrated that "exercise is medicine" for both prevention and treatment of major illnesses? The question cannot be answered by physics, chemistry, biology, engineering, or any other field of science, except psychology, because the answer lies in the operations of the human mind and the brain (Iso-Ahola, 2013, 2018). Similarly, it is the psychological scientists who are tackling the hardest problem of all problems in all of science (Gleiser, 2014): human consciousness. It does not matter even if knowledge is incomplete at the time when answers are provided. Much is known about antecedents of depression and how to treat its effects today, but more will be known tomorrow.

Even though psychological science does not seek to make precise predictions for individual behaviors and concomitant precise replications of them, this does not undermine its scientific status. Instead of precise predictions, psychological science aims to establish patterns and regularities to elucidate general human tendencies, and robust patterns likely explain recent replication successes (Klein et al., 2014; Camerer et al., 2018). However, patterns are just that, patterns, for which there are exceptions that make individual predictions imprecise. Even though specific behavioral predictions cannot be made from general human tendencies, knowing an increasing number of individuals' general tendencies can enhance the predictive power at the individual level. Despite the fact that "causality operates on single instances, not on populations whose members vary" (Cohen, 1994), statements regarding probabilistic causation can be made at the group level (Pearl and Mackenzie, 2019).

Thus, it is not possible to predict the time, the place, and the name of a particular individual to be involved in the next mass shooting, but factors influencing the probability of the phenomenon's occurrence can be determined. Even if the resultant information provides explanations in hindsight, cumulating knowledge can become increasingly useful in real life. Difficulties of predicting specific human acts and behaviors highlight why psychology is a complex and "hard" science.

However, inability to make precise predictions for specific behaviors and specific individuals' behaviors does not diminish the value of psychology as science, certainly no more than a recent failure of the Phillips Curve has been used to question economics as science. The Phillips Curve (i.e., a negative correlation between unemployment and inflation) has served as the basic tenet of economics for over 60 years and guided Federal Reserve in its policy decisions. Yet, the interactive relationships between wages, employment, and inflation are more complicated than any two-variable correlation. Similarly, an explanation of human behavior cannot be reduced to a psychological equivalence of the Phillips Curve. Complexity of human behavior calls for multi-variable theories to explicate stability and variability in cognition, affect, and behavior.

ROLE OF THEORY

Although science can be seen as a dynamic dance between theory, methods, and data (Boker and Martin, 2018), it is important to acknowledge that replications take a backseat to the most important work in the scientific process of discovering truth: theory-building, theory-elaboration of the latent structure of psychological phenomena, model generation, and "continuous model expansion" (Meehl, 1990; Gelman and Shalizi, 2013; Edelsbrunner and Dablander, 2020). Many believe that "scientific progress in psychology will likely depend upon the development of integrated models" (Grice, 2014, p. 22). Recent evidence, however, raises a note of caution about this process as complex models are less likely to true. Saylors and Trafimow's data (2020) showed that as the number of variables in the model increased, the probability that the model is true decreased rapidly.

Unquestionably, reality is complex if one counts all the hundreds of variables that could affect human affect, cognition, and behavior. In everyday life, though, people do not deliberate over 50 possible reasons why they should or should not go for a walk or run. Instead, they, at least regular exercisers, have delegated decision-making to cognitively less demanding operations of the non-conscious mind (Iso-Ahola, 2018), and this process should therefore be taken into account in theory-building. Sufficiently specified theory would indicate major variables or conditions under which effects could occur.

In general, scientists' task is to develop theories to explain how the universe functions and how the human mind operates, and for this, empirical feedback is needed. Empirical data can help clarify theories and contribute to the expansion of theoretical models, but they cannot turn psychological phenomena into yes-or-no particles whose existence is determined by replications

using arbitrary statistical criteria. Once a clear theoretical, logical, or mathematical case has been made for a phenomenon, be it gravity or cognitive dissonance, the phenomenon will not cease to exist as a deductive truth. Where would physics (and the world) be today without Einstein's theories? The Higgs boson particle was theorized (logically and mathematically) to exist in 1964 but not empirically verified until 2012. Did the particle not exist in the meantime?

Naturally, false, or even absurd, theories will be proposed from time to time, but they will quickly be dismissed when it is seen that they cannot stand logical, theoretical, or mathematical scrutiny. Granted, in some cases, such dismissals can be premature if theories have not been fully or sufficiently developed. This was evident when the initial idea of quantum mechanics (i.e., quantum entanglement of pairs of particles) was rejected by many physicists, including Einstein who called the entanglement "spooky action at a distance." Although he and his associates (Einstein et al., 1935) concluded that the quantum mechanics account of physical reality is incomplete, they left the door open by suggesting that "such a theory is possible." A critical development occurred when Bell published his mathematical "inequality" theorem in 1964, which enabled empirical testing of the two accounts of physical reality (quantum mechanics and Einstein's view). Bell's proposal inspired experimental work, but it took more than 50 years before the quantum mechanics explanation was confirmed beyond any reasonable doubt (Hensen et al., 2015; Yin et al., 2017).

There is an important lesson here for psychological science. Theories should not prematurely be dismissed, especially by single replication experiments or even meta-analyses, as empirical verification is a long-term process. Furthermore, given the complexity of psychological phenomena not as yes-or-no particles, empirical evidence is unlikely to be able to deal the final blow to a theory (Popper, 1959; Lakatos, 1976). In light of methodological difficulties of empirical verification, and the decades the process can take, it is illogical that single replication failures of many psychological effects have recently been accorded much more weight than numerous successful replications of the original (Doyen et al., 2012). Since single, direct replications (method replications), even if preregistered and involving many labs, are not any purer methodological demonstrations than the original, they cannot be final arbiters of various phenomena's existence or non-existence. If, however, replications are based on a diversity of methods and methodological improvements (phenomenon-focused replications), they will be useful for creating new knowledge. It should also be noted that failed replications can help advance science when they lead to revisions and reinventions of the original theory or hypothesis, as in the case of Inzlicht and Schmeichel's (2012) re-theorizing ego-depletion as a motivational mechanism rather than a resource depletion.

Theories are continually being refined, qualified, expanded, and their boundary parameters being established by empirical tests, as no theory is complete, especially in psychological science where time and context interactively cause considerable variation in behavior. Along with refinement and expansion, the basic idea of a well-developed theory provides the best provisional and propositional explanation for the nature of the universe or the operations of the human mind at the present time. Thus, even when its boundary conditions become better refined and established, the basic idea of cognitive dissonance will not disappear from the general psychological explanation of human behavior. In a similar vein, Newton's conception of gravity stands even after Einstein took Newton's theory to a different level by specifying the mechanism to explain gravity in space-time. The upshot is that in a bigger picture, theory and data are interconnected: "Data without theory is lame; theory without data is blind" (Gleiser, 2014). However, the recent elevation of replication into a special status in empirical examination has mistakenly led many to believe that scientific truth is a matter of the pinpoint statistical verification through replication. As argued, such a conclusion is groundless.

CONCLUSION

Scientific knowledge is temporary, dynamic, conditional, and relative to contexts in which an examined behavior occurs. As the scientific explanation is always provisional and propositional, no absolute truth exists, either in physics or psychology. Strictly speaking, as the physicist Ethan Siegel concluded, "scientific truth is a myth."

Time and context are inextricably interwoven in human behavior, with time creating different contexts and contexts creating different experiences of time. The resultant variability poses insurmountable problems for precise replications. Since exact replications are not possible, a replication study is just another empirical investigation in ongoing efforts to establish scientific truth. It does so by refining, qualifying, and expanding an earlier finding, but it cannot declare whether something exists permanently, or not at all. It is therefore more accurate to talk about temporary scientific knowledge than the presence or absence of scientific truth, regardless of claims made by replication researchers. Science builds knowledge incrementally and cumulatively and cannot therefore make categorical pronouncements on the existence or non-existence of a given truth by means of replications, especially single replications; nor can scientific truth be defined as statistical truth.

Psychological phenomena are essentially interactive effects, exerting their influence through many variables as opposed to stand-alone "main" effects. This complexity greatly increases the likelihood of replication failures. Moreover, because psychological phenomena are time- and context-bound, their causes are rarely exclusively either necessary or sufficient, and never both. This complex nature of causality of psychological effects creates major problems for demonstrations of the replicability of previous findings. While external (physical) conditions can be made to approximate those of the original study, however, internal conditions between the original and replication studies are never the same. That is, there is not a single report in the literature that would have shown a similar (much less identical) degree of both conscious and non-conscious processing by participants of the original and replication studies. In short, unless it is known what is in the minds of replication

study participants, one cannot be certain if the original study is being replicated.

The fact that exact replications are not possible does not mean that nothing can be known. Science is still "the best game in town" for creating new basic and applied knowledge for better understanding the universe and the human mind (Kerlinger, 1973). Even though scientific knowledge is limited and incomplete, it does not mean that psychological phenomena do not exist. Their existence, however, is not a matter of the pinpoint statistical verification through replication but primarily a function of a continual dynamic dance between theory, method, and data. In this process, phenomenon and constructive replications play an important role as long as they are continuous and iterative, avoiding declarations of the "basic existence" of phenomena and exercising constrains in making generalizations. Replications become constructive and useful for the advancement of science when they employ the same or similar method in different contexts and different methods in the same or similar contexts of human behavior on one hand, and when they lead to revisions and reformulations of existing hypotheses and theories on the other. In this process, emphasis is placed on phenomenon replication rather than method replication.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The author thanks John Bargh, Bradley Hatfield, and Matthew Miller as well as the reviewers for their most helpful and generous comments and suggestions on earlier versions.

REFERENCES

Abelson, R. (1985). A variance explanation paradox: when a little is a lot. *Psychol. Bull.* 97, 129–133. doi: 10.1037/0033-2909.97.1.129

Ackerman, P. L. (2014). Nonsense, common sense, and science of expert performance: talent and individual differences. *Intelligence* 45, 6–17. doi: 10.1016/j.intell.2013.04.009

Anderson, C., Bahnik, S., Barnett-Cowan, M., Bosco, F., Chandler, J., Chartier, C., et al. (2016). Response to comment on "Estimating the reproducibility of psychological science". *Science* 351:1037. doi: 10.1126/science.aad7243

Bargh, J. (1997). Reply to commentaries, in *Advances in Social Psychology*, ed. R. Wyer (Mahwah, NJ: Erlbaum), 231–246.

Bargh, J. (2014). The historical origins of priming as the preparation of behavioral responses: unconscious carryover and contextual influences of real-world importance, in *Understanding Priming Effects in Social Psychology*, ed. D. Molden (New York, NY: Guildford Press), 218–233.

Bargh, J. (2017). *Before You Know it*. New York, NY: Simon Schuster.

Baumeister, R. (2015). Conquer yourself, conquer the world. *Sci. Am.* 312, 61–65. doi: 10.1038/scientificamerican0415-60

Baumeister, R., Masicampo, E., and Vohs, K. (2011). Do conscious thoughts cause behavior? *Annu. Rev. Psychol.* 62, 331–361. doi: 10.1146/annurev.psych.093008.131126

Baumeister, R., Vohs, K., and Funder, D. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspect. Psychol. Sci.* 2, 396–403. doi: 10.1111/j.1745-6916.2007.00051.x

Bavel, J., Mende-Siedlecki, P., Brady, W., and Reinero, D. (2016). Contextual sensitivity in scientific reproducibility. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6454–6459. doi: 10.1073/pnas.1521897113

Belletier, C., Normand, A., and Huguet, P. (2019). Social-facilitation-and-impairment effects: from motivation to cognition and the social brain. *Curr. Direct. Psychol. Sci.* 28, 260–265. doi: 10.1177/0963721419829699

Berlyne, D. (1960). *Conflict, Arousal, and Curiosity*. New York, NY: McGraw-Hill.

Boker, S., and Martin, M. (2018). A conversation between theory, methods, and data. *Multiv. Behav. Res.* 53, 806–819. doi: 10.1080/00273171.2018.1437017

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* 62, 193–217. doi: 10.1037/h0047470

Bryan, C., Walton, G., Rogers, T., and Dweck, C. (2011). Motivating voter turnout by invoking the self. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12653–12656. doi: 10.1073/pnas.1103343108

Bryan, C., Yeager, D., and O'Brien, J. (2019). *Replicator Degrees of Freedom Allow Publication of Misleading "Failures to Replicate"*. Chicago, IL: University of Chicago.

Burger, J. (2009). Milgram replicated. would people still obey today? *Am. Psychol.* 64, 1–11. doi: 10.1037/a0010932

Bushman, B. (2018). Guns automatically prime aggressive thoughts, regardless of whether a "good guy" or "bad guy" holds the gun. *Soc. Psychol. Pers. Sci.* 9, 727–733. doi: 10.1177/1948550617722202

Camerer, C., Dreber, A., Holzmeister, F., Ho, T., Huber, J., Johannesson, M., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nat. Hum. Behav.* 2, 637–644. doi: 10.1038/s41562-018-0399-z

Carter, E., Kofler, L., Forster, D., and McCullough, M. (2015). A series of meta-analytic tests of tests of the depletion effect: self-control does not seem to rely on a limited resource. *J. Exper. Psychol. Gen.* 144, 796–815. doi: 10.1037/xge0000083

Casey, B., Somerville, L., Gotlib, I., Ayduk, O., Franklin, N., Askren, M., et al. (2011). Behavioral and neural correlates of delay of gratification 40 years later. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14998–15003. doi: 10.1073/pnas.1108561108

Chang, H. (1995). Circularity and reliability in measurement. *Perspect. Sci.* 3, 153–172.

Charlesworth, T. E., and Banaji, M. (2019). Patterns of implicit and explicit attitudes: I. Long-term change and stability from 2006 to 2016. *Psychol. Sci.* 30, 174–192. doi: 10.1177/0956797618813087

Cohen, J. (1994). The earth is round ($p < .05$). *Am. Psychol.* 49, 997–1003. doi: 10.1037/0003-066x.49.12.997

Crandall, C., and Sherman, J. (2015). On the scientific superiority of conceptual replications for scientific progress. *J. Exper. Soc. Psychol.* 66, 93–99. doi: 10.1016/j.jesp.2015.10.002

Cronbach, L., and Meehl, P. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957

Cumming, E. (2014). The new statistics: why and how. *Psychol. Sci.* 25, 7–29. doi: 10.1177/0956797613504966

De Vrieze, J. (2018). The metawars. *Science* 361, 1185–1188. doi: 10.1126/science.361.6408.1184

Dehaene, S. (2014). *Consciousness and the Brain*. New York, NY: Penguin.

Dijksterhuis, A., and Aarts, H. (2010). Goals, attention, and (un)consciousness. *Annu. Rev. Psychol.* 61, 467–490. doi: 10.1146/annurev.psych.093008.100445

Doliński, D., Grzyb, T., Folwarczny, M., Grzybała, P., Krzyszycha, K., Martynowska, K., et al. (2017). Would you deliver an electric shock in 2015? Obedience in the experimental paradigm developed by Stanley Milgram in the 50 years following the original studies. *Soc. Psychol. Pers. Sci.* 8, 927–933. doi: 10.1177/1948550617693060

Doyen, S., Klein, O., Pichon, C., and Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS One* 7:e29081. doi: 10.1371/journal.pone.0029081

Earp, B., and Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Front. Psychol.* 6:621. doi: 10.3389/fpsyg.2015.00621

Edelsbrunner, P., and Dablander, F. (2020). The psychometric modeling of scientific reasoning: a review and recommendations for future avenues. *Educ. Psychol. Rev.* 31, 1–34. doi: 10.1007/s10648-018-9455-5

Einstein, A., Podolsky, B., and Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* 47, 777–780. doi: 10.1103/physrev.47.777

Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramrakha, S., et al. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. *Psychol. Sci.* 31:792–806. doi: 10.1177/0956797620916786

Elms, A. (2009). Obedience lite. *Am. Psychol.* 64, 32–36. doi: 10.1037/a0014473

Enkavi, A., Eisenberg, I., Bissett, P., Mazza, G., MacKinnon, D., Marsch, L., et al. (2019). Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proc. Natl. Acad. Sci. U.S.A.* 116, 5472–5477. doi: 10.1073/pnas.1818430116

Ericsson, K., Krampe, R., and Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406. doi: 10.1037/0033-295x.100.3.363

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.

Foulk, T., Lanaj, K., and Krishnan, S. (2019). The virtuous cycle of daily motivation: effects of daily strivings on work behaviors, need satisfaction, and next-day strivings. *J. Appl. Psychol.* 104, 755–775. doi: 10.1037/apl0000385

Gelman, A., and Shalizi, C. (2013). Philosophy and the practice of bayesian statistics. *Br. J. Math. Statist. Psychol.* 66, 8–38. doi: 10.1111/j.2044-8317.2011.02037.x

Gerber, A., Huber, G., Biggers, D., and Hendry, D. (2016). A field experiment shows that subtle linguistic cues might not affect voter behavior. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7112–7117. doi: 10.1073/pnas.1513727113

Gilder, T., and Heerey, E. (2018). The role of experimenter belief in social priming. *Psychol. Sci.* 29, 403–417. doi: 10.1177/0956797617737128

Gleiser, M. (2014). *The Island of Knowledge*. New York, NY: Basic Books.

Goodman, S., Fanelli, D., and Ioannidis, J. (2016). What does reproducibility mean? *Sci. Transl. Med.* 8:341. doi: 10.1126/scitranslmed.aaf5027

Grice, J. (2014). Observation oriented modeling: preparing students for research in the 21st century. *Comprehen. Psychol.* 3, 1–27. doi: 10.2466/05.08.IT.3.3

Hagger, M., Chatzisarantis, N., and Zwienenberg, M. (2016). A multilab preregistered replication of the go depletion effect. *Perspect. Psychol. Sci.* 11, 546–573. doi: 10.1177/1745691616652873

Hanfstingl, B. (2019). Should we say goodbye to latent constructs to overcome replication crisis or should we take into account epistemological considerations? *Front. Psychol.* 10:1949. doi: 10.3389/fpsyg.2015.01949

Hensen, B., Bernien, H., Dréau, A. E., Reiserer, A., Kalb, N., Blok, M. S., et al. (2015). Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature* 526, 682–686. doi: 10.1038/nature15759

Hockey, G. (2011). A motivational control theory of cognitive fatigue, in *Cognitive Fatigue: Multidisciplinary Perspectives on Current Research and Future Applications*, ed. P. Ackerman (Washington, DC: American Psychological Association), 167–187. doi: 10.1037/12343-008

Hudson, N., Lucas, R., and Donnellan, M. (2017). Day-to-day affect is surprisingly stable: a 2-year longitudinal study of well-being. *Soc. Psychol. Pers. Sci.* 8, 45–54. doi: 10.1177/1948550616662129

Hussey, I., and Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Adv. Methods Pract. Psychol. Sci.* 3, 1–19. doi: 10.1037/pspa0000172

Inzlicht, M., and Schmeichel, B. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspect. Psychol. Sci.* 7, 450–463. doi: 10.1177/1745691612454134

Iso-Ahola, S. (2013). Exercise: why it is a challenge for both the nonconscious and conscious mind. *Rev. Gen. Psychol.* 17, 93–110. doi: 10.1037/a0030657

Iso-Ahola, S. (2017). Reproducibility in psychological science: when do psychological phenomena exist? *Front. Psychol.* 8:879. doi: 10.3389/fpsyg.2015.00879

Iso-Ahola, S. (2018). Human mind: both the cause and solution to the global pandemic of physical inactivity. *Intern. J. Public Health Res.* 6, 107–113.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York, NY: Farrar, Straus & Giroux.

Kerlinger, F. (1973). *Foundations of Behavioral Research*. New York, NY: Holt.

Klein, R., Ratliff, K., Vianello, M., Adams, R., Bahnik, S., Bernstein, M., et al. (2014). Investigating variation in replicability: a "Many Labs" replication project. *Soc. Psychol.* 45, 142–152. doi: 10.1027/1864-9335/a000178

Kok, B., Coffey, K. A., Cohn, M. A., Catalino, L. I., Vacharkulksemsuk, T., Algoe, S. B., et al. (2013). How positive emotions build physical health: perceived positive social connections account for the upward spiral between positive emotions and vagal tone. *Psychol. Sci.* 24, 1123–1132. doi: 10.1177/0956797612470827

Kuhn, T. (1962). *The Structure of Scientific Revolution*. Chicago, IL: University of Chicago Press.

Lakatos, I. (1976). Falsification and the methodology of scientific research programmes, in *Can Theories be Refuted?*, ed. S. Harding (Berlin: Springer), 205–259. doi: 10.1007/978-94-010-1863-0_14

LeBel, E., Berger, D., Campbell, L., and Loving, T. (2017). Falsifiability is not optional. *J. Pers. Soc. Psychol.* 113, 254–261. doi: 10.1037/pspi0000106

Lerner, J., Small, D., and Loewenstein, G. (2004). Heart strings and purse strings. *Psychol. Sci.* 15, 337–341. doi: 10.1111/j.0956-7976.2004.00679.x

Lieder, F., Griffiths, T., and Ming, H. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychol. Rev.* 125, 1–32. doi: 10.1037/rev0000074

Lin, C., Adolphs, R., and Alvarez, M. (2018). Inferring whether officials are corruptible from looking at their faces. *Psychol. Sci.* 29, 1807–1823. doi: 10.1177/0956797618788882

Loersch, C., and Payne, B. (2011). The situated inference model: an integrative account of the effects of primes on perception, behavior, and motivation. *Perspect. Psychol. Sci.* 6, 234–252. doi: 10.1177/1745691611406921

Lord, F., and Novick, M. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addition-Wesley.

Lykken, D. (1968). Statistical significance in psychological research. *Psychol. Bull.* 70, 151–159. doi: 10.1037/h0026141

Lykken, D. (1991). What's wrong with psychology, anyway?, in *Thinking Clearly About Psychology*, Vol. 1, eds D. Chicetti and W. Grove (Minneapolis, MN: University of Minnesota Press), 3–39.

MacNamara, B., Hambrick, D., and Oswald, F. (2014). Deliberate practice and performance in music, games, sports, education, and professions: a meta-analysis. *Psychol. Sci.* 25, 1608–1618. doi: 10.1177/0956797614535810

Malle, B. (2006). The actor-observer asymmetry in attributions: a (surprising) meta-analysis. *Psychol. Bull.* 132, 895–919. doi: 10.1037/0033-2909.132.6.895

Marsh, H., and Perry, C. (2005). Self-concept contributes to winning gold medals: causal ordering of self-concept and elite swimming performance. *J. Sport Exerc. Psychol.* 27, 71–91. doi: 10.1123/jsep.27.1.71

McFall, R. (1996). Making psychology incorruptible. *Appl. Prevent. Psychol.* 5, 9–15. doi: 10.1016/s0962-1849(96)80021-7

McShane, B., Bockenholt, U., and Hansen, K. (2020). Average power: a cautionary note. *Adv. Methods Pract. Psychol. Sci.* 3, 185–199. doi: 10.1177/2515245920902370

Meehl, P. (1967). Theory-testing in psychology and physics: a methodological paradox. *Phil. Sci.* 34, 103–115. doi: 10.1086/288135

Meehl, P. (1990). Appraising and amending theories: the strategy of Lakatosian defense and two principles that warrant it. *Psychol. Inq.* 1, 108–141. doi: 10.1207/s15327965pli0102_1

Meehl, P. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *J. Pers.* 60, 117–174. doi: 10.1111/j.1467-6494.1992.tb00269.x

Messick, S. (1989). Validity, in *Educational Measurement*, ed. R. Linn (Washington, DC: American Council on Education), 13–103.

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67, 371–378. doi: 10.1037/h0040525

Mischel, W. (1973). Toward a cognitive social learning reconceptualization of personality. *Psychol. Rev.* 80, 252–283. doi: 10.1037/h0035002

Mischel, W., Shoda, Y., and Peake, P. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *J. Pers. Soc. Psychol.* 54, 687–696. doi: 10.1037/0022-3514.54.4.687

Mischel, W., Shoda, Y., and Rodriques, M. (1989). Delay of gratification in children. *Science* 244, 933–938. doi: 10.1126/science.2658056

Murray, E. (1964). *Motivation and Emotion*. Englewood Cliffs, NJ: Prentice-Hall.

Nunnally, J. (1978). *Psychometric Theory*. New York, NY: McGraw-Hill.

Oppezzo, M., and Schwartz, D. (2014). Give your ideas some legs: the positive effect of walking on creative thinking. *J. Exper. Psychol.: Learning, Memory, and Cognition* 40, 1142–1152. doi: 10.1037/a0036577

Panek, R. (2020). A cosmic crisis. *Sci. Am.* 322, 32–37.

Pearl, J., and Mackenzie, D. (2019). *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books.

Platt, J. (1964). Strong inference. *Science* 146, 347–353. doi: 10.1126/science.146.3642.347

Popper, K. (1959). *The Logic of Scientific Discovery*. London: Hutchison.

Prentice, D., and Miller, D. (1992). When small effects are impressive. *Psychol. Bull.* 112, 160–164. doi: 10.1037/0033-2909.112.1.160

Pressman, S., and Cross, M. (2018). Moving beyond a one-size-fits-all view of positive affect in health research. *Curr. Direct. Psychol. Sci.* 27, 339–344. doi: 10.1177/0963721418760214

Ramscar, M. (2016). Learning and replicability of priming effects. *Curr. Opin. Psychol.* 12, 80–84. doi: 10.1016/j.copsyc.2016.07.001

Ramscar, M., Shaoul, C., and Baayen, R. H. (2015). *Why Many Priming Effects Don't (and Won't) Replicate: A Quantitative Analysis, Manuscript,* Germany: Tubingen University.

Rossi, J. (2013). Statistical power analysis, in *Handbook of Psychology: Vol 2. Research Methods in Psychology*, 2nd Edn, eds J. Schinka and W. Velicer (New York, NY: Wiley), 71–108.

Rubin, M. (2019). What type of Type I error? contrasting the neyman-pearson and fisherian approaches in the context of exact and direct replications. *Synthese* 196, 1–26. doi: 10.1007/s11229-019-02433-0

Ryan, L., Hatfield, C., and Hofstetter, M. (2002). Caffeine reduces time-of-day effects on memory performance in older adults. *Psychol. Sci.* 13, 68–71. doi: 10.1111/1467-9280.00412

Saylors, R., and Trafimow, D. (2020). Why the increasing use of complex causal models is a problem: on the danger sophisticated theoretical narratives pose to truth. *Organ. Res. Methods* 1094428119893452. (in press). doi: 10.1177/1094428119893452

Secrest, L. (1984). Reliability and validity, in *Research Methods in Clinical Psychology*, eds A. Bellack and M. Hersen (New York, NY: Pergamon), 24–54.

Shadish, W., Cook, T., and Campbell, T. (2002). *Experimental and Quasi-Experimental Designs For Generalized Causal Inference*. New York, NY: Houghton Mifflin.

Shamay-Tsoory, S., and Mendelsohn, A. (2019). Real-life neuroscience: an ecological approach to brain and behavior research. *Perspect. Psychol. Sci.* 14, 841–859. doi: 10.1177/1745691619856350

Sherman, J., and Rivers, A. (in press). There's nothing social about social priming: derailing the "train wreck". *Psychol. Inq.*

Sherry, D. (2011). Thermoscopes, thermometers, and the foundations of measurement. *Stud. His. Philos. Sci.* 42, 509–524. doi: 10.1016/j.shpsa.2011.07.001

Simmons, J., Nelson, L., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632

Snipper, E., Jeronimus, B., Aan Het Rot, M., Bos, E., de Jonge, P., and Wichers, M. (2018). The reciprocity of prosocial behavior and positive affect in daily life. *J. Pers.* 86, 139–146. doi: 10.1111/jopy.12299

Stroebe, W., and Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspect. Psychol. Sci.* 9, 59–71. doi: 10.1177/1745691613514450

Trafimow, D. (2003). Hypothesis testing and theory evaluation at the boundaries. *Psychol. Rev.* 110, 526–535. doi: 10.1037/0033-295x.110.3.526

Trafimow, D. (2019). A frequentist alternative to significance testing, p-values, and confidence intervals. *Econometrics* 7:26. doi: 10.3390/econometrics7020026

Tversky, A., and Kahneman, D. (1971). Belief in the law of small numbers. *Psychol. Bull.* 76, 105–110. doi: 10.1037/h0031322

Vuletich, H., and Payne, K. (2019). Stability and change in implicit bias. *Psychol. Sci.* 30, 854–862. doi: 10.1177/0956797619844270

Watts, T., Duncan, G., and Quan, H. (2018). Revisiting the marshmallow test: a conceptual replication investigating links between early delay of gratification and later outcomes. *Psychol. Sci.* 29, 1159–1177. doi: 10.1177/0956797618761661

Wegner, D. (1994). Ironic processes of mental control. *Psychol. Rev.* 101, 34–53. doi: 10.1037/0033-295x.101.1.34

Witt, W., Linkenauer, S., and Profitt, D. (2012). Get me out of this slump! Visual illusions improve sports performance. *Psychol. Sci.* 23, 397–399. doi: 10.1177/0956797611428810

Witte, E., and Zender, F. (2018). Data replication matters to an underpowered study, but replicated hypothesis corroboration counts. *Behav. Brain Sci.* 8:e156. doi: 10.1017/S0140525X18000924

Wood, W., Quinn, J., and Kashy, D. (2002). Habits in everyday life: thought, emotion, and action. *J. Pers. Soc. Psychol.* 33, 1281–1297. doi: 10.1037/0022-3514.83.6.1281

Yin, J., Cao, Y., Li, Y.-H., Liao, S.-K., Zhang, L., Ren, J.-G., et al. (2017). Satellite-based entanglement distribution over 1200 kilometers. *Science* 356, 1140–1144. doi: 10.1126/science.aan3211

Zhao, J., Al-Aidroos, N., and Turk-Browne, N. (2013). Attention is spontaneously biased toward regularities. *Psychol. Sci.* 24, 667–677. doi: 10.1177/0956797612460407

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

REVIEW
published: 17 September 2020
doi: 10.3389/fpsyg.2020.545959

Does Postmodernism Really Entail a Disregard for the Truth? Similarities and Differences in Postmodern and Critical Rationalist Conceptualizations of Truth, Progress, and Empirical Research Methods

*Peter Holtz**

IWM Leibniz-Institut für Wissensmedien, Tübingen, Germany

OPEN ACCESS

Edited by:
Peter Adriaan Edelsbrunner, ETH Zürich, Switzerland

Reviewed by:
Luca Tateo, University of Oslo, Norway
Barbara Hanfstingl, Alpen-Adria-Universität Klagenfurt, Austria

***Correspondence:**
Peter Holtz
p.holtz@iwm-tuebingen.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *26 March 2020*
Accepted: *25 August 2020*
Published: *17 September 2020*

Citation:
Holtz P (2020) Does Postmodernism Really Entail a Disregard for the Truth? Similarities and Differences in Postmodern and Critical Rationalist Conceptualizations of Truth, Progress, and Empirical Research Methods. Front. Psychol. 11:545959. doi: 10.3389/fpsyg.2020.545959

Within this article, I will compare postmodernist and critical rationalist conceptualizations of epistemological key concepts such as truth, progress, and research methods. An analysis of Gergen's program for a postmodern psychology shows that a naïve positivist understanding of truth is clearly incompatible with his postmodernist approach, whereas a correctly understood falsificationist use of truth as a guiding ideal may not be. However, postmodernists are often content with a diversity of voices as the endpoint of scientific activities, whereas critical rationalists such as Popper would put more emphasis on attempts to reach a common understanding. The differences between critical rationalists such as Popper and Deutsch and postmodernists such as Gergen are more complicated when it comes to conceptualizations of progress: whereas, postmodernists do not deny the existence of some forms of progress such as technological innovation, they argue that the modernist grand narrative, which views Western culture and the corresponding technological revolutions as being equal to epistemological progress and societal and political progress *per se*, has become untenable. Debates on possible negative consequences of modern technology are one example of evidence for this. Here, critical rationalists tend to engage in a legitimization discourse, sensu Lyotard, and to defend Western culture with all its deficiencies as a necessary precondition for evolutionary epistemic as well as societal and political progress, although they would agree with large parts of the postmodern critique of modernism. Postmodernists and critical rationalists would both agree that psychology as a field would benefit greatly, among other things, from a transition from a methods-oriented approach to scientific knowledge to a more problem-oriented approach, and from less methodological dogmatism. Taken together, postmodernism and critical rationalism may not be as irreconcilable as it may seem at first glance.

Keywords: postmodernism, critical rationalism, epistemology, philosophy of science, Kenneth Gergen, Karl Popper, David Deutsch, Jean François Lyotard

INTRODUCTION

There are numerous examples of philosophers as well as psychologists and other intellectuals warning of the destructive powers of the obscure specter of postmodernism (cf. Jauß, 1983), which to them is apparently haunting not only modern psychology, but also all of contemporary science and society. The psychologist and internet personality Jordan B. Peterson defines postmodernism on his homepage as follows (Peterson, undated):

> Postmodernism is essentially the claim that (1) since there are an innumerable number of ways in which the world can be interpreted and perceived (and those are tightly associated) then (2) no canonical manner of interpretation can be reliably derived.
>
> That's the fundamental claim. An immediate secondary claim (and this is where the Marxism emerges) is something like "since no canonical manner of interpretation can be reliably derived, all interpretation variants are best interpreted as the struggle for different forms of power."

In a similar vein, theoretical physicist and critical rationalist philosopher David Deutsch describes postmodernism as follows (Deutsch, 2011, p. 314):

> One currently influential philosophical movement goes under various names, such as postmodernism, deconstructionism, and structuralism [sic!], depending on historical details that are unimportant here. It claims that because all ideas, including scientific theories, are conjectural and impossible to justify, they are essentially arbitrary: they are no more than stories, known in this context as 'narratives'. Mixing extreme cultural relativism with other forms of anti-realism, it regards objective truth and falsity, as well as reality and knowledge of reality, as mere conventional forms of words that stand for an idea's being endorsed by a designated group of people such as an elite or consensus, or by a fashion or other arbitrary authority. And it regards science and the Enlightenment as no more than one such fashion, and the objective knowledge claimed by science as an arrogant cultural conceit.

Psychologist and linguist David Pinker recently made the following statement in an interview with the British newspaper *The Guardian* (Anthony, 2018):

> If scientific beliefs are just a particular culture's mythology, how come we can cure smallpox and get to the moon, and traditional cultures cannot? And if truth is just socially constructed, would you say that climate change is a myth? It's the same with moral values. If moral values are nothing but cultural customs, would you agree that our disapproval of slavery or racial discrimination or the oppression of women is just a western fancy?

One difficulty in disentangling all this criticism of some diffuse notion of postmodernism is that those who came to be known as the founding fathers and mothers of postmodernism, such as Foucault, Derrida, Lacan, and Irigaray, had never used the term themselves (Wilterdink, 2002, p. 197). Hence, in order to assess the veracity of the claims by Peterson, Deutsch, Pinker, and others, one must first narrow down the notion of postmodernism that is to be analyzed.

Undoubtedly, Lyotard (1979/1984) has largely contributed to the popularization of the term postmodernism in its current meaning with his book on the *postmodern condition*, which – according to Google Scholar – has, as of now, been cited more than 30,000 times in scientific texts. For the field of psychology, the explicit agenda of Gergen (1990) for a postmodern psychology seems to be a seminal text that can serve as a point of reference for the postmodernist movement in the field of psychology.

Based mainly on Gergen's conceptualization of postmodernism (see also Gergen, 1994, 2001) and to some degree "classic" of Lyotard (1979/1984), I will begin by discussing implications for the field of psychology, to lead into a discussion of differences and similarities between postmodernist and critical rationalist conceptualizations of truth, progress, and empirical research methods.

POSTMODERNISM

Modernism

The term *postmodernism* originated in fields as different as philosophy, architecture, and literary theory in the early twentieth century (Wilterdink, 2002). The common element among different conceptualizations of postmodernism seems to be the idea that another era, *modernism*, had reached its end and was to be replaced by a yet-to-be-named new epoch. On a side note, I would like to add to the existing accounts of the origins of postmodernism that in the early twentieth century, the term *hypermodernism* (Tartakower, 1924) was fiercely debated in the world of chess: previously, a modernist movement mirroring the scientific method in trying to discover abstract rules for identifying promising and detrimental positional features and general strategic patterns had replaced an earlier, more romantic approach to chess that focused mostly on mating attacks (e.g., Tarrasch, 1912). In the early twentieth century, a group of grandmasters such as Reti and Nimzovitch had discovered that there were exceptions to the modernist "laws" of good chess that could be used to skillfully outmaneuver dogmatic modernists. However, no hypermodernist at the time denied that the modernist principles were valid for most chess positions. Rather, the object of criticism was the dogmatic and oversimplifying tone of modernism.

In a similar vein, the postmodern movement in the sciences should also be understood first and foremost as a countermovement against some extreme form of modernism (see also Lyotard, 1979/1984, pp. 11–14). Gergen (1990) describes psychology itself as a through-and-through modernist attempt to replace earlier humanities-based approaches to understanding

the human condition with a profoundly empirical approach based on the "scientific method" of the natural sciences (for a discussion of the naive positivistic way in which the scientific method is usually understood in psychology, see Holtz and Monnerjahn, 2017). This "modernist romance" (p. 25) is, according to Gergen, characterized by four overarching presuppositions:

1. Basic subject matter: there is something that can be known – be it behavior and its causes or internal processes, such as thoughts or memories and the like. The important thing is that although we do not (yet) know everything, there is agreement (apart from some "localized conflicts"; Gergen, 1990, p. 25) on the subject matter that is to be known and on the fact that the subject matter can be known.
2. Universal properties: modernists assume that they can derive by inductive reasoning from single observations certain abstract, general, and time-invariant laws of nature that explain a class of phenomena that are related to the basic subject matter.
3. Empirical method: in line with what Gergen calls a "logical empiricist" (p. 26) approach (without providing further references), modernists believe that through strict application of the scientific method they can gain true knowledge about the universal properties of the basic subject matter. They believe they can thus prevent the "entry of ideology, values or passions into the description and explanation of relevant phenomena" (p. 26).
4. Research as progressive: by applying the scientific method, scientists can abandon false beliefs and "move toward the establishment of reliable, value neutral truths about our designated segment of the objective world" (p. 26).

Gergen continues by providing a very rough account of various forms of criticism of the logical empiricist's (or logical positivist's) philosophy supposedly underlying psychology's modernism, represented by philosophers such as Quine, Popper, and Kuhn. It seems pretty clear that Gergen is here and elsewhere not referring to certain philosophers from the logical positivist tradition in particular, but to psychology's (or more specifically social psychology's) positivism-influenced epistemological tradition (see also Pettigrew, 1991; Holtz and Monnerjahn, 2017).

In the following paragraph, I will outline how and why the modernist dream of positive knowledge about a segment of the objective world has in my opinion pretty clearly reached its limits in the present time. I will use a technology-focused approach similar to Lyotard (1979/1984) in discussing the internet as a research object.

The Internet as a Thoroughly Postmodern Phenomenon

If we apply those four presuppositions of modernism to the internet, the modernist will quickly run into some problems: the basic subject matter, for example, the internet and specific internet-related activities, came into being only two and a half decades ago. At least those who remember the internet from its early days will certainly agree that what is there and what can be done now does not have much to do with the humble beginnings in the 1990s. And even worse, I assume that none of us will doubt that "the internet" or whatever it is going to evolve into will, 25 years from now, not have much in common with what we are experiencing now.

Furthermore, the internet already offers so many facets and services that the experiences possible within this environment can hardly be subsumed under the same overarching theoretical principles (see also Orben, 2020). That is, of course, based on the assumption that the internet does not change us and is only shaped by us (or other things such as economic interests) driven by psychological principles that also exist outside and beyond the internet. As Gergen (1973) had argued elsewhere, not only is the idea that the environment does not influence the psyche untenable, it is also highly probable, given that most psychological articles end with a concluding statement emphasizing the societal impact of the research reported, that our psychological theorizing about the world influences the very world we analyze. At the very least, we either must admit that our research is without societal consequences or we must abandon the idea of general time-invariant laws that govern human's minds and behavior.

In view of the internet's ever-changing subject matter, it is also untenable to think that experiments or any other empirical research method can guarantee the discovery of true statements and the abandonment of false statements. Much time has passed since vision of Allport (1924) that in the "near future," psychology will have discovered the basic psychic processes that lead to complex social structures, so that "social objects" (such as groups, societies, norms, cultural codes…) would finally lose any explicatory function. Postmodernists such as Gergen demand that this dream of an explanation of the social based upon processes within individuals be finally put to rest for good. In contrast, human individuality, including processes such as thinking and arguing, as well as behavior itself, can only be understood against the background of a cultural fabric that underlies all psychic processes.

Within the internet, it also becomes obvious how knowledge has become a good that is traded just like other goods and that acquires its worth from supply and demand. In earlier days, the production of new knowledge and the administration of existing knowledge had been by and large the privilege of a dedicated class within society such as first priests and monks and then scientists at universities who were employed by and acted in the interest of the respective religious and secular authorities. Knowledge was in these days more of an "end in itself" (Lyotard, 1979/1984, p. 5) than it is today. Lyotard (1979/1984) goes to great lengths to explain how knowledge has already become (as of the 1980s) and will become even more in the future (i.e., now) *the* "principle force of production" (p. 16) that drives our globalized economy. Hence, all claims to knowledge are necessarily also of economic and political interest. Any claims to "value-neutrality" can be nothing more than a sales argument in a globalized knowledge economy.

> Knowledge is and will be produced in order to be sold,
> it is and will be consumed in order to be valorized in a

> new production: in both cases, the goal is exchange. Knowledge ceases to be an end in itself, it loses its "use-value" (p. 15/16).

Lastly, there is the question of progress. Several arguments can be made against the modernist idea of continuous progress. On the one hand, changes in the subject matter can make something that at time T1 was progress obsolete at T2. Let us imagine that Kraut et al. (1998) "internet paradox," the idea that the internet as a communication technology eventually makes users lonelier, was true in the 1990s, but untrue in the 2000s. We can also easily imagine that changes in technology could make the statement true again in the 2010s or 2020s. So, what is progress and how can we know that we have made progress?

The question of progress can also be tackled from a very different perspective: what *is* progress? Does progress mean better control over people's behavior on the internet? Catering to psychological needs with better technologies? Better health and less loneliness? What if abandoning the internet along with any related research was the best progress achievable, as some technology pessimists seem to imply? And who decides what progress is and on which basis? These are some of the core questions postmodernists ask modernists.

Gergen's Vision for a Postmodern Psychology

As we have seen, using the scientific method to identify time and context-invariant psychological laws of nature is not feasible from a postmodern point of view. So, what remains to be done for the postmodern psychologist? Already in his earlier works, Gergen (1978) had introduced the ideal of creating *generative theories*, that is, theories that allow for challenging established assumptions about the world and for exploring alternative lifestyles and behavioral patterns. To Gergen (1990), postmodernism opens up a whole new realm of possibilities: whereas psychologists in the modern age were assigned the finally doomed Sisyphus task of finding natural laws where there may be only historically and culturally shaped transient and volatile patterns, they can now, for example, contribute instead to the creation of a better world. In the modernist vision of science, the scientist is a passive observer and analyst. In the postmodern vision, scientists can use their locally and temporally limited knowledge to explore possible worlds, and they can try to answer the question how a change for the better can be facilitated – from a certain perspective at a certain point in time within a given socio-historical context.

CRITICAL RATIONALISM

In a Nutshell

In the following paragraph, I will attempt to summarize the central "mantra" of critical rationalism as concisely as possible. More elaborate accounts of critical rationalism, and the rather misleading way in which it is most often characterized in psychological texts, have been provided elsewhere (e.g., Holtz, 2016; Holtz and Monnerjahn, 2017; Holtz and Odağ, 2018). Gergen's criticism is targeting primarily modernist psychology and not the "natural sciences," hence I will refer as a primary point of reference to essay of Popper (1976/1969) *the logic of the social sciences* whenever possible.

The central question of Popper (1959/2002), at least s *the logic of scientific discovery* (LoSD), was how we can at the same time admit that all our knowledge is fallible and still rationally justify a belief in the possibility of a growth of knowledge. Popper began his scientific career at a point in time when some of the pillars of physics – the showcase project of modernity – had just been scattered by the "Einsteinian revolution." Hence, one could no longer ignore the possibility that even our most highly valued intellectual tenets, such as Newtonian mechanics, could possibly turn out to be wrong and be replaced by better theories at any point in time. Still, to Popper, it would be just silly to insist that there is no progress, when at the same time, science, technology, and society had just begun to evolve at an unprecedented pace.

Maybe his most important insight was that the concept of truth in an absolute sense is not needed to believe in progress: if a new theory explains everything that an old theory could explain (for example, but not only, in the sense of making correct predictions) and explains additional phenomena that the old theory could not explain – that is progress, simple as that. Science should accordingly be done in a way that facilitates exactly this kind of progress: scientists should make bold predictions that can easily be shown to not correspond with certain observations (falsification), so that one can as easily as possible identify ways to improve upon them (Popper, 1970). The belief in an absolute truth can hence even easily hinder progress, since one cannot improve upon a supposedly absolute truth.

Critical Rationalism and Modernism/Positivism

Critical rationalism was a response to positivist/inductivist approaches, such as the logical empiricism of the "Vienna circle" that aimed at defining verification criteria, that is, procedures that allow scientists do discern true from false statements. Ideally, scientific knowledge should only (or predominantly) be based upon verified elementary statements (e.g., Reichenbach, 1938; Carnap, 1967/1928). However, as the famous US-American A.J. Ayers logical positivist famously pointed out in a TV-interview with British philosopher Bryan Magee in 1978, logical positivism failed and finally fell into disfavor with epistemologists, because no viable verification procedure could ever be identified (the full interview can be found at PhilosophyOverdose, undated).

Popper also argued against the idea that the social sciences should just "copy" the methods of the natural sciences, such as experiments (Popper, 1976/1969, p. 90). Just like the concept of an absolute truth can hinder progress, methods that appear to guarantee true knowledge can also forestall progress, for example, if methods that are meant to discover discrepancies between expectations and observations are used to "prove" theories. Popper calls this uncritical copying of research methods from the natural sciences "scientism" (ibid.) and "misguided naturalism" (p. 91).

The way in which methods such as experiments are used in psychology to create evidence in favor of theories was maybe most sharply criticized from a falsificationist perspective by another critical rationalist, Imre Lakatos:

> After reading Meehl (1967) and Lykken (1968) one wonders whether the function of statistical techniques in the social sciences is not primarily to provide a machinery for producing phoney corroborations and thereby a semblance of 'scientific progress' where, in fact, there is nothing but an increase in pseudo-intellectual garbage (Lakatos, 1978, p. 88).

What social scientists should attempt to copy instead from the "older" natural sciences is the critical approach that can be found often, but not always, among leading physicists and chemists: the possibility of a growth of knowledge critically depends on the willingness of the protagonists within scientific discourse to expose their ideas to criticism and to admit it when discrepancies between their expectations and observations emerge (falsification). They also have to be willing to change their beliefs if someone can offer a better explanation (see above) for the phenomena to be explained. In the section on "progress" below, we will discuss in more concrete terms different interpretations of the implications of a critical rationalist epistemology for the social sciences.

Two aspects which are related to our previous discussion of postmodernism should be noted here: first, Popper readily admits that several mutually incompatible accounts of an event can exist; this seems to me to correspond in many aspects to concept of Lyotard (1979/1984) of a narrative. Second, proponents of different narratives can and should still attempt to exchange views and to – whenever this is possible – reach a consensual position, just like speakers of different languages can at least try to reach a common understanding (Popper, 1978). They make the attempt although there is no guarantee that they will come to a common understanding and although the understanding they reach will certainly likely be less than perfect. Hence, apart from a critical stance, scientific progress also depends critically on the willingness to communicate in a consensus-oriented way.

In the following paragraphs, I will try to summarize and to directly compare the postmodern (mainly sensu Gergen, 1990) and the critical rationalist stance toward three pivotal epistemological concepts: notions of truth, epistemic progress, and the role and function of methods in scientific inquiry.

TRUTH

Throughout all of Popper's works, truth in an absolute or metaphysical sense must be discerned from individual cases where assumptions about the world (theories or hypotheses) apparently correspond to observations. The absolute truth status of scientific hypotheses or theories can never be clarified once and for all. This is because, for example, explanatory hypotheses and theories both refer to an infinite class of phenomena (e.g., every X under condition Y). Hence, even if all observations had so far corresponded to our theories, one can never be sure that all future observations will do so as well (this problem is sometimes called the Humean problem of induction, e.g., Popper, 1959/2002, p. 5).

Additionally, to Popper, there are no theory-free observations: all "facts" that we derive from our senses should be understood as answers to questions that we formulated beforehand or as tentative solutions to problems that we have tried to solve. Hence, all insights from empirical research are necessarily preliminary:

> Knowledge does not start from perceptions or observations or the collection of data or facts, but it starts, rather, from problems. One might say: No knowledge without problems; but also, no problems without knowledge. But this means that knowledge starts from the tension between knowledge and ignorance (Popper, 1976/1969, p. 88).

On a side note, it should be noted here that also a falsification of a theory is itself fallible: we could discover at any point, for example, that we ignored some boundary conditions or auxiliary hypotheses which led us to falsely believe that our theory was false.

Popper was also fully aware that the questions we ask or the problems we attempt to solve are culturally bounded: "*Ninth thesis*: A so-called scientific subject is merely a conglomerate of problems and attempted solutions, demarcated in an artificial way. What really exists are problems and solutions and scientific traditions (Popper, 1976/1969, p. 92)."

However, Popper also frequently warned of the dangers of the "malaise of existentialism" (Popper, 1976/1969, p. 104) that could result from an erroneous interpretation of the insight that all our knowledge is fallible: the fact that we cannot know anything for certain and that all insights are to some degree culturally bounded does not allow for the conclusion that one error is just as bad or good as another and that researchers cannot at least attempt to find increasingly better solutions for problems.

Immediately after his "ninth thesis" (see above), Popper tells the story of an interdisciplinary meeting on the future of humanism that he once attended, in which a cultural anthropologist took part as well. In the following paragraph, Popper mocks (in the voice of the anthropologist) the anthropologist's relativist stance in being unwilling to discuss the arguments that the participants brought forward with regard to the topic of the meeting:

> While arguments or reasons make an impression on you, as participants in a discussion, what interests us is the fact that through such means you can mutually impress and influence each other; and also of course the symptoms of this influence. We are concerned with concepts such as emphasis, hesitation, intervention, and concession. We are actually not concerned with the factual content of the discussion but only with the role which the various participants are playing: with the dramatic interplay as such. As to the so-called arguments, they are of course only one aspect of verbal behaviour

> and not more important than the other aspects (Popper, 1976/1969, p. 94; emphasis as in the original).

It is important to note that what Popper criticizes first and foremost is what he views as the cultural anthropologist's arrogance in assuming that his understanding of the situation is more objective than the other participants' viewpoints. He also criticizes the anthropologist's claim to be able to discern safely between "objective observations" of "verbalizations" and other forms of behavior and objectively invalid pseudo-arguments of the participants that only serve some obscure political purpose. Popper certainly dislikes the unwillingness of the anthropologists to at least attempt to solve the problem at hand and the anthropologist's cynical ridiculing of any attempt to make progress and to reach a mutual understanding.

To my understanding, Popper does not criticize the *questions* that postmodernists ask of other scientists (and social scientists in particular), such as the question about societal power structures that are at play in academic discourse as well. However, Popper does dislike the fact that postmodernists often (at least, apparently, according to his experience) seem to think that they have an *objective* or otherwise privileged answer to these questions. If a postmodern criticism of a debate such as the one outlined above could and would be formulated as "testable hypotheses" in the sense of debatable and criticizable statements, such criticism could constitute an important element within a critical rationalist attempt to gain an increasingly refined understanding of societal phenomena.

In the last paragraphs, I introduced a term that may be a bit more difficult to understand than Popper's concept of truth: Popper's concept of *objectivity*, which played a pivotal role in his later scientific work from the 1960s onward (see, e.g., Popper, 1976/1969). How can there be "objective knowledge" if we cannot have truth? The only answer to this question can be that objectivity is as much of an unreachable regulative ideal as is truth. The opposite of objectivity here would be subjectivity, in the sense that a statement is only comprehensible for me or that an insight about the world makes sense for me, but I cannot communicate it successfully to others (as in Wittgenstein's concept of a *private language* that cannot be understood by others; Wittgenstein, 1958/1953, $259, p. 92 f.). In this sense, objectivity seems to equal intersubjectivity: a statement can only be objective to the degree that critical, but well-meaning (in the sense that they are not cynical and that they are genuinely interested in finding a solution to the problems that are under discussion) participants will reach a common understanding of the statement and a consensus about the (of course tentative) truth status of the statement at a given point in time within a certain socio-historical context. To Popper, scientists in their discussions should strive for increasing objectivity just as much as they are supposed to strive for the truth. But claims of "absolute" objectivity make just as little sense as do claims for an "absolute" truth. An important tool for achieving increasing objectivity is to Popper the application of formal logical principles.

However, things get complicated through Popper's frequent attempts to belittle the roles that, for example, societal structures, the socio-historical context or the socio-cultural embeddedness of individual researchers play in academic discourse – and particularly in discourse in the social sciences: "Such minor details as, for instance, the social or ideological habitat of the researcher, tend to be eliminated in the long run; although admittedly they always play a part in the short run" (Popper, 1976/1969, p. 96). It seems that Popper's horror of a postmodernist relativist skepticism that renders any attempts to solve problems futile makes him sometimes talk as if critical scientists could grasp the objective (not-subjective) aspects of a problem or of proposed solutions in an absolute way. His frequent recourse to formal logical arguments and the possibility of deducting hypotheses from theories can intensify the impression that objectivity can be grasped, whereas truth remains an unreachable ideal. However, such a reading of Popper is, in my opinion, self-contradictory and untenable. In consequence, although Popper frequently expressed dislike toward thinkers such as Foucault (see, e.g., Horgan, 2018), his ideas may have been closer to postmodern thinkers than he was aware of himself (see also Holtz, 2016).

It is interesting to compare Popper's criticism of relativism with criticism of "objective knowledge" of Gergen (1994) in his response to criticism of Smith (1994) his agenda for a postmodern psychology:

> Consider the ideal of objective knowledge. In psychology, as in other sciences, the claim to 'objective knowledge' operates as a conversational trump. It disregards or denigrates all hands not dealt in these terms (e.g., evidence, measurement, reliability). Any views not based on scientific tenets—for example, those of sundry religions, political action groups, ethnicities, genders, cultures—can be dismissed as folk beliefs—or more pejoratively, as value-biased, superstitious, or despotic. In terms of its relational implications, 'science talk' is thus as totalizing as that of the demagogy that science has sought to replace (Gergen, 1994, p. 413).

But is this the objective knowledge that Popper had in mind? I would think that Gergen is criticizing here the positivist/modernist psychologists' claim to have access to *actual* objective knowledge, and not so much the use of objectivity as a guiding ideal. Popper would probably agree that claims to objective knowledge are problematic and can indeed be easily abused to justify discrimination and other forms of power games. Popper (1945) criticized exactly this misuse of claims to objective knowledge, for example, in his "open society." However, it must be noted that whereas Gergen does not want to draw a boundary between science and other societal institutions such as religion, Popper, particularly in his early works, attempted to differentiate between science and non-science on the basis of the falsifiability = criticizability of its tenets (e.g., Popper, 1959/2002, p. 10 ff.). We will come back to this question in the paragraphs on the conceptualizations of progress and the role of empirical research methods.

We finally arrive at the question as to whether Popper's concept of truth as a regulative ideal can be reconciled with Lyotard's and Gergen's postmodernist approaches. I would argue

that they are indeed compatible (see also Holtz, 2016; Holtz and Odağ, 2018). If we take, for example, statement of Gergen (1990) that instead of researchers who are just "objectifying the taken-for-granted assumptions of the culture" (p. 33), we need scholars who are "willing to be audacious, to break the barriers of common sense by offering new forms of theory, of interpretation, of intelligibility" (ibid.), this could very well also be a sentence from one of Popper's later works, as long as these audacious scholars also display humility and the willingness to expose their new theories to criticism.

The same is true for demand of Gergen (1990) that psychologists should focus more on societal problems and on creating a better world: "Required, then, is a form of professional investment in which the scholar attempts to de-objectify the existing realities, to demonstrate their social and historical embeddedness, and to explore their implications for social life." To my understanding, this statement resembles quite closely Popper's proposal to improve societies by means of small-scale societal experiments in his "open society" (Popper, 1945).

A non-naive reading of Popper's use of terms such as truth and objectivity as guiding ideals (be it the understanding Popper intended or not) is indeed compatible with a not completely "radical" postmodernist or constructivism (see also Gadenne, 2008). I mean this in the sense that it is not regarded as outright impossible to reach some form of a common understanding and consensus among well-meaning participants in discourse. Of course, such a consensus is fallible, and it occurs against a certain socio-historical background.

In constructivism, intersubjectivity is needed as well as means of reaching a consensual understanding of social constructions. Hence, attempts at mutual understanding are necessarily at the core of any postmodern research agenda. However, critical rationalism would ask the participants in a discourse to go beyond mutual understanding in that they are also asked to attempt not only to understand each others' constructions of the world, but also to reach a common understanding with regard to, for example, problems, that are to be solved and the assessment of proposed solutions to these problems. Postmodernism sensu Gergen, in contrast, seems to merely aim at acknowledging and giving voice to different social constructions and world views; critical rationalism also endorses diversity, but attempts should be made at reaching some form of common sense whenever that is possible. Here, postmodernists will most likely be afraid that in the attempt to find common ground and common sense privileged groups will be likely to normatively enforce their world views upon less privileged groups. To the critical rationalist, this is a valid concern, but giving up attempts at reaching a mutual understanding in the sense of consensually negated assessments of problems and proposed solutions would mean to give up any chance for societal or epistemological progress, and this is not an option for the critical rationalist.

PROGRESS

Progress in the form of replacing theories with better theories is a central concept in critical rationalist thinking. Although to Popper all knowledge is preliminary, he would maintain that one can hardly deny that there has been some form of progress in science that mirrors the increasing complexity of life forms caused by (at least at the level of individuals) seemingly chaotic attempts at propagating genes (see, e.g., Popper, 1971/1961). Deutsch (2011) proposed a maybe even more radical evolutionary epistemology than Popper's in describing the acquisition of knowledge as an epiphenomenon of life's evolution: just like life forms evolved from attempts to solve problems such as survival and from the fact that successful adaptations to environments manifested themselves in transmittable DNA structures, our knowledge is the result of millennia of more or less successful attempts at problem solving and attempts to codify the outcomes of these trials in cultural products, such as human language and among others, the cultural tradition called science:

> Both in science and in biological evolution, evolutionary success depends on the creation and survival of objective knowledge, which in biology is called adaptation. That is, the ability of a theory or gene to survive in a niche is not a haphazard function of its structure but depends on whether enough true and useful information about the niche is implicitly or explicitly encoded there (Deutsch, 1998, p. 48).

Of course, knowledge is not accumulated in a monotonous way, in the sense that true statements about the world are kept whereas wrong statements are dismissed (for a criticism of this positivistic notion of a cumulative growth of knowledge see Kuhn, 1962, p. 169 ff.). There can always be the kinds of temporary setbacks Kuhn (1962) described in his "structure" (e.g., p. 111 ff.). Totalitarian societal structures that prohibit asking critical questions and trying out new solutions for problems can even forestall any progress whatsoever for some time. However, according to Deutsch (2011, p. 64 ff.), we can observe since the enlightenment an increased speed in the acquisition of knowledge (in the sense of a cultural product) that goes along with societal and political developments, such as increasing freedom and democracy in Western societies. Deutsch (2011) frequently dismisses calls for a change toward more "sustainability" as attempts to restore the anti-progressive totalitarian order of earlier and darker epochs (e.g., p. 434 f.). Deutsch does not deny the existence of problems such as global warming and environmental pollution (e.g., Deutsch, 2011, p. 440 f.), but to him going backward does not constitute a viable attempt at solving these problems. The critical rationalist idea of progress critically relies on a certain degree of optimism, in the sense of belief that at least some of the problems we face can indeed be solved.

Is this now the "grand narrative of progress" (Gergen, 1990, p. 30) to which we cannot return anymore according to Lyotard (1979/1984, p. 60), and which Gergen attempts to "demystify" (Gergen, 1990, p. 33)? What Gergen seems to dislike most about the Western grand narrative is that it silences other voices such as, for example, the voice of less privileged members of a society and the voice of non-Western societies altogether. To some degree,

critical rationalists would agree that no voice must be silenced as long as it does not itself demand the silencing of other voices in a totalitarian sense (see Popper, 1945). However, Popper and Deutsch would certainly argue that there are rational, and to some degree objective, reasons to prefer (of course, in a still culturally bounded and fallible way) a Western democratic society with all its deficiencies over, for example, a totalitarian theocracy that is ruled by brutish religious zealots. I would think that postmodernists such as Gergen do not have an easy answer to this argument, since the improvement of life-worlds is one of the central elements of Gergen's outline of a postmodern psychology:

> In the postmodern vein, we find that all languages—even that of the research psychologist—can enter the culture and be used by people to justify, separate, control, and castigate. In effect, for the psychologist there is no escaping matters of moral and political consequence. This being the case, not only is it irresponsible to avoid deliberations on the good, but psychologists should be encouraged to add their voices to the culture's dialogues of "ought." In many cases this may mean political advocacy—championing causes that one believes good for the culture; in others it may mean culture critique—condemning movements or policies that seem inimical to human welfare (Gergen, 1994, p. 414/415).

Postmodernists criticize the somewhat authoritarian (see also Feyerabend and Oberheim, 2011) modernist narrative of a cumulative growth of knowledge (see above), but critical rationalists do so as well. It should also be noted that neither Lyotard nor Gergen rule out the possibility of some forms of progress, such as technological innovation (e.g., Gergen, 1990, p. 31). However, to Gergen, the main contribution of psychology to society is not so much technology, for example, in the form of new and innovative ways of measuring psychological properties and approaches to the treatment of psychiatric diseases. Instead, the main function of psychology, at least in a modern Western democracy, could be to explore different forms of seeing the world and of being in the world: psychology can bring into the open voices that are the most often marginalized within the mainstream societal discourse. However, I do not think that a critical rationalist would object to this endeavor, as long as it does not lead to the cynical refusal to believe in the possibility of any form of consensus and progress. It is of course also understandable that a physicist such as Deutsch focuses in his epistemology first and foremost on what Gergen calls technological innovation and not so much on social and societal issues.

When it comes to societal progress or attempts at creating a better world in general, idea of small-scale societal experiments of Popper (1945) seems to be not so much different from Gergen's concept of generative theories (e.g., Gergen, 1978): both concepts have in common that trying out new ways of living and of organizing social life is crucial for societal progress. It is not the case that scientists first create "objective" or "true" knowledge which is then eventually applied, for example, by politicians; first, new solutions to existing problems must be found and tried out. One difference is that critical rationalists such as Popper or Deutsch emphasize more strongly that the resulting ideas from concepts such as generative theory or small-scale societal experiments have to be evaluated critically:

> This growth, this self-transcendence, has a rational side and a non-rational side. The creation of new ideas, of new theories, is partly non-rational. It is a matter of what is called 'intuition' or 'imagination'. But intuition is fallible, as is everything human. Intuition must be controlled through rational criticism, which is the most important product of human language. This control through criticism is the rational aspect of the growth of knowledge and of our personal growth. It is one of the three most important things that make us human. The other two are compassion, and the consciousness of our fallibility (Popper, 1978, p. 167).

However, the idea that a critical evaluation of the outcomes of generative theories is needed can be found in Gergen's writings as well: "As assumptions are sustained or rejected, social life may be altered in ways that may be judged 'good' or 'bad' from some standpoint" (Gergen, 1978, p. 1356). Again, the two approaches may differ from each other more in terms of prioritization than in terms of substance.

To my understanding, postmodernists and critical rationalists have in common that they first and foremost argue against different forms of totalitarianism and that they both encourage the expression of deviating opinions and different voices. However, critical rationalists seem to abhor primarily the specter of a cynical relativism that renders futile any attempts at mutual understanding and at facilitating a change for the better, postmodernists are more concerned about a dogmatic positivistic scientific culture that considers itself superior and to some degree infallible and that has a tendency to silence critics and alternative approaches. It seems to be that both approaches share similar concerns, but their protagonists may have had personal experiences with different forms of dogmatism, which may have led to different sets of fears and concerns.

METHODS

To Gergen (1990), a postmodernist turn would also have methodological implications: as Gergen (2001) discussed elsewhere, laboratory experiments can most often be regarded as "degraded data" and "myopic" (p. 810) from the perspective of a postmodern psychologist who is interested in exploring the depth and richness of culturally bounded discourse patterns and behavioral repertoires. Qualitative methods of inquiry may often be better suited to explore different constructions of reality and the effects of, for example, certain psychological theories or viewpoints on the emergence of corresponding life-worlds. However, Gergen also mentions the potential of "classical" social psychological laboratory experiments

(e.g., Asch, 1956; Milgram, 1974) to incite "public discussion on issues of political and societal significance" (p. 808).

Overall, postmodernist researchers seem to be free in choosing any research method they like as long as they are aware that all scientific inquiries are in the end acts of communication within a culturally bound tradition and a system of meaning making. Hence, no claims for an objective or absolute truth can be deduced from any research method. What research methods can do for the postmodernist is that they can aid her in understanding – from her culturally bounded position – the plurality of culturally embedded psychological realities and the ways in which such realities can change under certain conditions.

To Popper (1976/1969), it would of course be foolish to think of certain research methods as pathways to the truth or to exclude certain research methods on ideological grounds. To the critical rationalist, the role and function of empirical research methods are to allow for criticism of theories in the form of giving them a chance to fail. For example, Popper's whole philosophy is not based on empirical studies, but on thought experiments as well as on formal logic and other rational arguments. Every scientific discipline will need a range of different methods to expose their respective theories to criticism. Hence, one can very well imagine that different research methods are more or less useful in different areas of the so-called natural and social sciences, which can be no more than loosely defined traditions to the critical rationalist.

However, truth and objectivity should be the guiding ideals in the critical rationalists' choice of methods. What does this mean and what would be the implications for the social scientist? Research methods should help us to overcome "psychologism" (e.g., Popper, 1959/2002, p. 7), that is the idea that insights that appear to be true to the beholder cannot be shared with others in a form that allows for mutual understanding and criticism. It should be noted that this kind of objectivity in the sense of intersubjectivity is also one of the main goals of literally all approaches within the wide field of qualitative research methods (see also Holtz and Odağ, 2018).

Popper himself certainly favored different kinds of research methods in the social sciences. In the "logic of the social sciences," Popper (1976/1969, p. 103) briefly recommends the kind of "situational analysis" that is used in economics as a possible approach for the social sciences: here, the researcher assumes that human beings behave (more or less) rationally and tries to identify the situational factors under which a certain type of behavior would be rational. It seems to follow that also systematic deviations from the assumption of rationality, for example, in the form of "biases" (e.g., Tversky and Kahneman, 1974) can be analyzed, and the results of these analyses can be used to improve situational analyses.

From a postmodern perspective, this approach to social scientific research can be critziced, for example, because it may tend to ignore the socio-cultural boundedness of social practices and the constructedness of rationality itself. Maybe a critical rationalist would respond that as long as such criticism is presented in a way that it yields debatable or empirically testable assumptions, it can perfectly well be reconciled with a critical rationalist approach to research. Personally, I think that Popper's apparent predilection for a more homo oeconomicus orientied approach in the social sciences is certainly not the only way in which critical rationalist social scientific research can play out. If we look, for example, at ideas of Pettigrew (1991) on a critical rationalist social psychology, he is envisioning a stronger unity between the different branches within social psychology such as experimental social psychology and a more qualitatively oriented "symbolic interactionist" (p. 13) approach. He also suggests trying to build bridges to more humanities-based approaches in sociology. Hence, a critical rationalist approach to psychology could also be imagined as a more inclusive enterprise bridging the existing gaps between qualitative and quantitative approaches above and beyond the mere "mixing" of methods (cf. Holtz and Odağ, 2018). Just like in the previous paragraphs on truth and progress, Popper's (or any other philosopher's) personal preferences do not count much in view of the question how their philosophical approaches can and should be interpreted.

Taken together, postmodernists as well as critical rationalists take a pragmatic stance when it comes to research methods. Postmodernists tend to prefer research methods that allow for the reconstruction of different forms of discourse, whereas most critical rationalists may value approaches that allow for increasingly objective arguments away from intersubjectivity. Still, there do not seem to be any irreconcilable differences here.

IMPLICATIONS FOR THE FIELD OF PSYCHOLOGY

It is fairly easy to find a common enemy of postmodernists and critical rationalists: modernist psychologists who insist that only experiments (or other research methods that are supposedly borrowed from the hard natural sciences) guarantee true knowledge. By insisting this, they immunize their theories against criticism since any objection that does not result from experiments can be easily devalued as being unscientific. This is particularly worrying since experiments were almost eclusively used over the last decades in the field of psychology to "prove" or to "sell" theories, and not to subject them to severe tests as means of improving on them (Holtz, 2020). Thus, modernist psychologists unfairly make their worldview privileged, excluding other voices from the perspective of the postmodernists. Such approaches can easily be misused as propaganda tools in socio-political power struggles.

Both postmodernists and critical rationalists would ask scientists to be bold, to try out new ways, and to bring diverging opinions to the front. Science is not about being timid and hiding behind pompous technical language (Billig, 2013) or haughty and complicated research methods. Science is – or should be – an adventure (cf. Willig, 2001), and scientists should have the audacity to ask questions that have never been asked before and to try out new solutions to old problems. No knowledge, no theory, and no empirical research is sacrosanct; everything can and must be questioned at any time. I do not see much of a difference between both approaches here.

Both approaches also encourage social scientists to think of themselves as parts of the world and not as passive and objective observers. Popper writes in the "logic of the social sciences":

> Serious practical problems, such as the problems of poverty, of illiteracy, of political suppression or of uncertainty concerning legal rights were important starting points for research in the social sciences. Yet these practical problems led to speculation, to theorizing and thus to theoretical problems. In all cases, without exception, it is the character and the quality of the problem—and also of course the boldness and originality of the suggested solution—which determine the value, or the lack of value, of a scientific achievement (Popper, 1976/1969, p. 89).

However, to Popper, progress in science is always progress in the forms of the development of increasingly "better" theories. The solving of societal problems can probably only be a byproduct of social scientific research. Gergen (1990) goes to great length in his texts to avoid any notion of this kind of progress: all knowledge that might result from any scientific activity can only be understood within its own cultural tradition and cannot claim superiority over other forms of knowledge that resulted from other cultural traditions. However, in view of Gergen's optimism that politically and societally a change for the better can be achieved, I find it difficult to believe that he really rules out such positive developments for the realm of scientific knowledge. In the following statement, for example, Gergen is discussing the merits of the newly emerging field of theoretical psychology:

> The point of criticism should not be that of terminating traditions or practices but of helping them to evolve in ways that more fully integrate the voices of the discipline and of its constituents and contribute to the intellectual resources of the world (Gergen, 2001, p. 809).

To me, it is difficult to think of a successful contribution to the intellectual resources of the world without employing some concept of better and worse contributions. However, Gergen would probably maintain that this concept itself is culturally and temporally bounded. Popper and Deutsch would agree that no claim to knowledge can be "objective," but they would probably use formal logic and other "rational" arguments to support their belief in the possibility of getting closer to the truth by means of constant trial and error. Here, it would be up to the postmodernists to answer the question whether their relativism with regard to a growth of knowledge is primarily meant to be an attack against naïve positivist modernist psychologists or whether they really believe "in the depth of their hearts" that there cannot be progress, or that objectivity and truth are not ideals that are worth striving for.

On the other hand, critical rationalists such as Popper and Deutsch will have to justify their optimism in believing that striving for truth and objectivity will finally lead to primarily positive consequences. At the same time, critical rationalists have to be aware of the dangers that even well-meaning attempts to contribute to the creation of a better world can turn out to be disastrous and that attempts at reaching a mutual understanding among different voices can unwillingly lead to the establishment of cultural hegemony. Although Deutsch, in particular, sometimes speaks as if the superiority of post-enlightenment Western culture was an "objective fact" (see the quote in the section Introduction), the concepts of fallibility and error correction do of course apply here as well.

CONCLUSION

In the end, there are many commonalities between critical rationalist and postmodern approaches: a dislike for modernist arrogance, an emphasis on boldness and innovation, and a struggle against totalitarian attempts to oppress voices in a discourse. Both approaches value attempts by social scientists to address real world problems and to challenge dogmas and established world views. Critical rationalists and postmodernists are both aware that all our knowledge is temporally and culturally bounded, insofar as we can only perceive the world from a certain discursive formation (Foucault, 1969/1972) or that we can only ask questions and receive empirical answers to these questions from a certain point in time within given cultural structures.

The differences between critical rationalists and postmodernists boil down to differences in preferences and predilections: whereas postmodernists sometimes challenge arrogant modernist conceptualizations in a cynical way, critical rationalists prefer to propagate optimism with regard to the question as to whether there are at least some problems that we, as scientists, may be able to solve in a tentative and preliminary way. Whereas postmodernists are mostly afraid of dogmatic empiricists that hide their own political agenda behind claims for objective truth, critical rationalists are first and foremost weary of zealous postmodernists who themselves hide *their* political agenda behind their own claims of objectivity. Whereas postmodernists value a plurality of different voices, critical rationalists hope for a consensual resolution of conflicts by means of relying of increasingly objective arguments.

On a personal note, I do not think that any of these differences are beyond reconciliation. Both approaches value open and free discussions above everything else, and this fact alone should provide common ground for attempts at increasing mutual understanding.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

FUNDING

The present research was funded by the Leibniz Association, Germany [Leibniz Competition 2018, funding line "Collaborative Excellence", project SALIENT (K68/2017)].

REFERENCES

Allport, F. H. (1924). *Social psychology*. Boston: Houghton.

Anthony, A. (2018). Steven pinker: 'The way to deal with pollution is not to rail against consumption.' Available at: https://www.theguardian.com/science/2018/feb/11/steven-pinker-enlightenment-now-interview-inequality-consumption-environment (Accessed February 2, 2020).

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.: Gen. Appl.* 70, 1–70. doi: 10.1037/h0093718

Billig, M. (2013). *Learn to write badly: How to succeed in the social sciences*. Cambridge: Cambridge university pess.

Carnap, R. (1967/1928). *The logical structure of the world (German original "Der logische Aufbau der welt" 1928)*. Chicago: Open court.

Deutsch, D. (1998). *The fabric of reality*. Westminster: Penguin UK.

Deutsch, D. (2011). *The beginning of infinity: Explanations that transform the world*. Westminster: Penguin UK.

Feyerabend, P., and Oberheim, E. (2011). *The tyranny of science*. Cambridge: Polity Press.

Foucault, M. (1969/1972). *The archeology of knowledge* (translated by A. M. Sheridan Smith; French original 1969). New York: Pantheon.

Gadenne, V. (2008). The construction of realism. *Constr. Found.* 3, 153–159.

Gergen, K. J. (1973). Social psychology as history. *J. Pers. Soc. Psychol.* 26, 309–320. doi: 10.1037/h0034436

Gergen, K. J. (1978). Toward generative theory. *J. Pers. Soc. Psychol.* 36, 1344–1360. doi: 10.1037/0022-3514.36.11.1344

Gergen, K. J. (1990). Toward a postmodern psychology. *Humanist. Psychol.* 18, 23–34. doi: 10.1080/08873267.1990.9976874

Gergen, K. J. (1994). Exploring the postmodern: perils or potentials? *Am. Psychol.* 49, 412–416. doi: 10.1037/0003-066X.49.5.412

Gergen, K. J. (2001). Psychological science in a postmodern context. *Am. Psychol.* 56, 803–813. doi: 10.1037/0003-066X.56.10.803

Holtz, P. (2016). How Popper's 'Three Worlds Theory' resembles Moscovici's 'Social Representations Theory' but why Moscovici's 'Social Psychology of Science' still differs from Popper's 'Critical Approach. *PSR* 25, 13.1–13.24.

Holtz, P. (2020). Two questions to foster critical thinking in the field of psychology: are there any reasons to expect a different outcome, and what are the consequences if we don't find what we were looking for? *Meta Psychol.* 4:894. https://doi.org/10.15626/MP.2018.984

Holtz, P., and Monnerjahn, P. (2017). Falsificationism is not just 'potential' falsifiability, but requires 'actual' falsification: social psychology, critical rationalism, and progress in science. *J. Theory Soc. Behav.* 47, 348–362. doi: 10.1111/jtsb.12134

Holtz, P., and Odağ, Ö. (2018). Popper was not a positivist: why critical rationalism could be an epistemology for qualitative as well as quantitative social scientific research. *Qual. Res. Psychol.* 17, 541–564. doi: 10.1080/14780887.2018.1447622

Horgan, J. (2018). The paradox of Karl Popper. Available at: https://blogs.scientificamerican.com/cross-check/the-paradox-of-karl-popper/ (Accessed September 01, 2020).

Jauß, H. R. (1983). *Zur historischen Genese der Scheidung von Fiktion Realität*. München: Fink.

Kraut, R., Patterson, M., Lundmark, V., Kiesler, S., Mukophadhyay, T., and Scherlis, W. (1998). Internet paradox: a social technology that reduces social involvement and psychological well-being? *Am. Psychol.* 53, 1017–1031. doi: 10.1037//0003-066x.53.9.1017

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (1978). "Falsification and the methodology of scientific research programmes" in *The methodology of scientific research programmes*. eds. J. Worrall and G. Currie (Cambridge: Cambridge university press), 8–93.

Lykken, D. T. (1968). Statistical significance in psychological research. *Psychol. Bull.* 70, 151–159. doi: 10.1037/h0026141

Lyotard, J. F. (1979/1984). *The postmodern condition: A report on knowledge* (translated by Geoff Bennington and Brian Massumi; French original 1979). *Vol. 10*. Minneapolis: University of Minnesota Press.

Meehl, P. E. (1967). Theory-testing in psychology and physics: a methodological paradox. *Philos. Sci.* 34, 103–115. doi: 10.1086/288135

Milgram, S. (1974). *Obedience to authority: An experimental view*. London: Tavistock Publications.

Orben, A. (2020). Teenagers, screens and social media: a narrative review of reviews and key studies. *Soc. Psychiatry Psychiatr. Epidemiol.* 55, 407–414. doi: 10.1007/s00127-019-01825-4

Peterson, J. B. (undated). Postmodernism: definition and critique (with a few comments on its relationship with Marxism). Available at: https://www.jordanbpeterson.com/philosophy/postmodernism-definition-and-critique-with-a-few-comments-on-its-relationship-with-marxism/ (Accessed February 2, 2020).

Pettigrew, T. F. (1991). "Toward unity and bold theory: popperian suggestions for two persistent problems of social psychology" in *The future of social psychology*. eds. C. W. Stephan, W. G. Stephan and T. Pettigrew (New York: Springer), 13–27.

Philosophy Overdose (undated). A. J. Ayer on logical positivism & its legacy [video file]. Available at: https://www.youtube.com/watch?v=VAqicPNeKYo (Accessed February 2, 2020).

Popper, K. R. (1945). *The open society and its enemies*. London: Routledge.

Popper, K. R. (1959/2002). *The logic of scientific discovery* (German original 1935). London: Routledge.

Popper, K. R. (1970). "Normal science and its dangers" in *Criticism and the growth of knowledge*. eds. I. Lakatos and A. Musgrave (Cambridge: Cambridge University press), 51–58.

Popper, K. R. (ed.) (1971/1961). "Evolution and the tree of knowledge (based on the Herbert Spencer lecture, delivered in Oxford in 1961)" in *Objective knowledge: An evolutionary approach* (Oxford: Clarendon), 255–284.

Popper, K. R. (1976/1969). "The logic of the social sciences" in *The positivist dispute in German sociology* (translated by G. Adey and D. Frisby; German original 1969). eds. T. Adorno, H. Albert, R. Dahrendorf, J. Habermas, H. Pilot and K. R. Popper (New York: Harper Torchbooks), 90–104.

Popper, K. R. (1978). *Three worlds*. Ann Arbor, MI: University of Michigan Press.

Reichenbach, H. (1938). *Experience and prediction: An analysis of the foundations and the structure of knowledge*. Chicago: University of Chicago press.

Smith, M. B. (1994). Selfhood at risk: postmodern perils and the perils of postmodernism. *Am. Psychol.* 49, 405–411. doi: 10.1037/0003-066X.49.5.405

Tarrasch, S. (1912). *Die moderne schachpartie. Kritische studien über mehr als 200 ausgewählte meisterpartien der letzten zwölf jahre mit besonderer berücksichtigung der eröffnungen (The modern chess game…)*. Zürich: Olms.

Tartakower, S. (1924). *Die hypermoderne Schachpartie [the hypermodern chess game]*. Zürich: Olms.

Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science* 185, 1124–1131. doi: 10.1126/science.185.4157.1124

Willig, C. (2001). *Introducing qualitative research in psychology: Adventures in theory and method*. New York: McGraw Hill.

Wilterdink, N. A. (2002). The sociogenesis of postmodernism. *Eur. J. Sociol./Arch. Eur. Sociol.* 43, 190–216. doi: 10.1017/S0003975602001078

Wittgenstein, L. (1958/1953). *Philosophical investigations. 2nd Edn*. Oxford, UK: Blackwell.

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 18 September 2020
doi: 10.3389/fpsyg.2020.562992

Why Most Research Findings About Psi Are False: The Replicability Crisis, the Psi Paradox and the Myth of Sisyphus

Thomas Rabeyron[1,2]*

[1]Université de Lorraine, Interpsy, Nancy, France, [2]University of Edinburgh, KPU, Edinburgh, United Kingdom

The replicability crisis in psychology has been influenced by the results of nine experiments conducted by Bem (2011) and presented as supporting the existence of precognition. In this paper, we hope to show how the debate concerning these experiments could be an opportunity to develop original thinking about psychology and replicability. After a few preliminary remarks about psi and scientific epistemology, we examine how psi results lead to a paradox which questions how appropriate the scientific method is to psi research. This paradox highlights a problem in the way experiments are conducted in psi research and its potential consequence on mainstream research in psychology. Two classical experiments – the Ganzfeld protocol and the Bem studies – are then analyzed in order to illustrate this paradox and its consequences. Mainstream research is also addressed in the broader context of the replication crisis, decline effect and questionable research practices. Several perspectives for future research are proposed in conclusion and underline the heuristic value of psi studies for psychology.

Keywords: psi, precognition, replicability crisis, pre-registration, methodology

OPEN ACCESS

Edited by:
Barbara Hanfstingl, Alpen-Adria-Universität Klagenfurt, Austria

Reviewed by:
Walter Erich Manfred Von Lucadou, Independent Researcher, Freiburg, Germany
Marc Wittmann, Institute for Frontier Areas of Psychology and Mental Health (IGPP), Germany

***Correspondence:**
Thomas Rabeyron
thomas.rabeyron@gmail.com

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *17 May 2020*
Accepted: *28 August 2020*
Published: *18 September 2020*

Citation:
Rabeyron T (2020) Why Most Research Findings About Psi Are False: The Replicability Crisis, the Psi Paradox and the Myth of Sisyphus. Front. Psychol. 11:562992. doi: 10.3389/fpsyg.2020.562992

REPLICABILITY CRISIS AND PSI RESEARCH

The replicability crisis has been illustrated by the results of nine experiments conducted by Bem (2011) and reported in the *Journal of Personality and Social Psychology* as supporting the existence of precognition. As Romero (2019) explains "although the finding persuaded very few scientists, the controversy engendered mistrust in the ways psychologists conduct their experiments because Bem used procedures and statistical tools that many social psychologists use" (p. 3). Indeed, if Bem was able to demonstrate the existence of precognition – and given that precognition cannot exist for a lot of psychologists (Reber and Alcock, 2020) – did he show unwittingly that something was profoundly wrong in the way experiments are conducted in the field of psychology (Wiggins and Chrisopherson, 2019)? Many relevant papers have been published since Bem's initial publication (Pashler and Harris, 2012; Savalei and Dunn, 2015) about the replicability crisis, Bayesian statistics (Witte and Zenker, 2017), and questionable research practices (QRPs; Wagenmakers et al., 2011; Bierman et al., 2016). In the present paper, we would like to suggest that this debate could be an opportunity to develop original thinking about psychology and replicability. In this regard, we will show that the Bem studies

are not an isolated "accident," but are actually inserted in a long tradition of research which tries to deal with complex epistemological problems concerning the nature of reality and human consciousness. Specifically, we will argue that the controversies about the existence of psi could be highly informative about psychology and consciousness studies.

Psi research can be considered as a subfield of consciousness studies concerned with interactions between individuals and their environment that transcends the ordinary constraints of space-time (Bem and Honorton, 1994; Radin, 2006; Irwin and Watt, 2007; Cardeña, 2018). Different lines of research have been developed for more than a century to tackle psi using experimental research (Rhine et al., 1966), spontaneous cases (Rhine-Feather and Schmicker, 2005), clinical cases (Rabeyron, 2020), selected participants[1] (Méheust, 1999; Braude, 2007), and applications (Schwartz, 2001, 2007; May et al., 2018). Several meta-analyses of studies conducted under controlled conditions examine precognitive dreams (*es* = 0.14; Sherwood and Roe, 2003), telepathy (*es* = 0.14; Storm et al., 2010), and presentiment (*es* = 0.21; Mossbridge et al., 2012) and have demonstrated statistically significant effects. Bem (2011) research about precognition[2] published in the JPSP is thus not isolated. It can be considered as the logical of evolution of previous psi research.

While these results support the existence of consistent anomalous experience/behavior that has been labeled "psi," there is currently no consensus in the scientific community concerning their interpretation and two main positions have emerged so far. The "skeptics" suppose that they are the consequences of errors, bias, and different forms of QRPs (Alcock, 2003; Alcock et al., 2003; Hyman, 2010; Wiseman, 2010; Wagenmakers et al., 2011; Reber and Alcock, 2020). The "proponents" argue that these results prove the existence of psi beyond reasonable doubt and that new research should move on to the analysis of psi processes rather than yet more attempts to prove its existence (Radin, 2006; Cardeña et al., 2015; Cardeña, 2018). This absence of consensus is related to the difficulty of drawing firm conclusions from the results of psi research. Indeed, they represent an anomaly (Rao and Palmer, 1987) because there is currently no scientific model – based on physical or biology principles – to explain such interactions even if they exist (Kuhn, 1962)[3]. Nevertheless, which ever explanation is correct, the results of psi research may be informative for the wider psychological sciences (Schooler et al., 2018). Indeed, they lead to two opposites but very heuristic hypothesis: (a) within this domain of research, which has been conducted by hundreds of researchers whose critical efforts span over a century, the researchers have either been fraudulent or have been fooled, even when using the most reliable tools of scientific research and (b) psi exists and human consciousness can interact with its environment beyond the usual boundaries of space and time. This paper will explore these two hypotheses and their consequences for psi research and psychology in general.

WHAT IF… PSI *REALLY* EXISTS? THE PSI PARADOX

History of science has shown many examples of phenomena that were observed in a reliable manner but which were rejected by the scientific community because they were not explainable at the time of their observation (e.g., meteorites, heartbeat, etc.). Consequently, it might be wise to be careful when considering anomalous results of the psi variety, especially knowing that 22 Nobel Prize winners, leading scientists, and figures of the intellectual life have reported such experiences and took a position in favor of their existence (Méheust, 1999)[4]. Thus, if we suppose – at least for a moment – that there are enough elements to take *seriously* the hypothesis that the (b) option is true, that is, psi *really* exists, what are the consequences of such an assumption? Could it make sense of other observations in the field of psychology and other scientific domains? And does it change the way scientific research should be conducted?

The psi studies, viewed as a whole, suggest that a kind of "direct" interaction (conscious or unconscious) between individual humans and their environment is possible. This interaction concerns events or objects situated at a distance in space and time (Mossbridge and Radin, 2018). It can take many forms (gut feeling, behavior, mental representation, etc.) with different intensities (e.g., a small or a strong emotion). It can be perceptive (from the environment to the person) or projective (from the person to the environment). It can be associated with a transfer of information or, rather, something that *looks like* a transfer of information (Lucadou et al., 2007). It emerges more easily during altered states of consciousness (Storm et al., 2010), is more pronounced for some persons (Schlitz and Honorton, 1992), and tends to emerge during or after traumatic events (Rabeyron and Loose, 2015).

[1]This was the classical approach of psychical research at the beginning of the 20th century, and it has progressively disappeared in favor of studies with unselected participants (Méheust, 1999). This is problematic because the data suggest that the original method is probably more relevant (Schlitz and Honorton, 1992; May et al., 2018). The approach relying on unselected participants yields small effect sizes that can only be shown at a statistical level, and is thus easily criticized and less convincing. On the contrary, working with "gifted" participants tends to yield larger effects (Eisenbud, 1966; Braude, 2007). Three examples of this kind of experiments, conducted in control conditions, are: an historical case, with Alexis Didier who would have been able a number time to read a word and sentences in a book at a distance (Méheust, 2003); the pictures were produced by Ted Serios on a polaroid camera while he was locked in a Faraday Cage (Eisenbud, 1966); and the Pearce-Pratt experiment in which Pearce has been able to guess at a distance the right figures (among five possibilities) 25 times in a row (Rhine et al., 1966). The argument usually proposed by skeptics to explain such results is that both the experimenter and the participant were cheating (Hansel, 1961; Palmer, 2016).

[2]A meta-analysis about this effect has also reached a significant statistical effect size (*es* = 0.14; Bem et al., 2015).

[3]Psi research is not unique in this regard. It can be argued that many effects in psychology have no underlying explanation from a biological or a physical point of view. But these "classical" effects are easily accepted because they do not question the knowledge from these domains, which is the case of psi. It is indeed not usual that psychology say something about reality that could contradict sciences studying matter and living organisms.

[4]See https://psi-encyclopedia.spr.ac.uk/articles/eminent-people-interested-psi

This direct interaction is more generally associated with subjective paranormal experiences (e.g., near death experiences and out of body experiences) even if they do not overlap entirely (Rabeyron et al., 2018). Moreover, even if the role of consciousness in psi processes is not well understood, it is likely that attention (McMoneagle and May, 2014), memory (Carpenter, 2012), creativity (Holt, 2012), personality traits (Thalbourne, 2000), belief (Lawrence, 1993), and psychodynamic aspects (Rabeyron and Loose, 2015) are components of it.

These results from psi research construct a specific paradox that is the crux of the present paper. This paradox would lead to the conclusion that most research findings in psi research are *false* (or inappropriate) but not for the reasons usually supposed by the skeptics. We need first to recall that scientific research is based on the principle that, in the main, the researcher (the observer) is separated from, or independent of, the dependent variable. The researcher tries not to affect the result of an experiment so that what is observed varies with the independent variable being tested and is independent of the researcher's thoughts, intentions, belief, or disbelief. This assumption is assumed to permit other scientists to demonstrate the same result under the same conditions (or not, as the case may be). This is the logical and epistemological frame in which scientific research is usually conducted. This model works very well and has produced much reliable scientific knowledge and technological progress. Even if this is a simplistic vision of the way scientific research progresses, as shown by epistemology of science (Feyerabend, 1975; Chalmers, 1979), such a representation of scientific research is nevertheless a useful principle, which guides scientists to good practice.

However, if a direct interaction (in the sense described above) between a person and their environment is possible, this principle too could influence the outcome of experiments purporting to use the scientific principles, because there *could* be a direct interaction between the scientists (the observer) and their object of study (the observed). Thus, if psi exists, the problem is the following: *an advertent or inadvertent "direct" interaction between the researcher and the object of study could be possible. This destroys the conditions necessary for the convincing scientific demonstration of psi itself.*

This leads to the following paradoxs: (1) if the existence of psi is proven in a classical scientific setting, then it demonstrates retroactively that this setting is inappropriate because psi implies that there is no clear "cut" between the observer and the observed. (2) So, if this assumption of the scientific setting is inappropriate (because of psi), then this setting cannot be used to prove the existence of psi. But, then, as psi cannot be proven, the scientific setting itself appears as being still appropriate! Consequently, we can try to use the scientific setting to prove the existence of psi but then we are now again at (1), which leads to (2), and it logically follows an infinite paradoxical loop between (1) and (2). It shows more generally that the principle on which a scientific experiment is based to study psi – the ontological separation between the observer and the observed – would be erroneous. Consequently, any scientific knowledge in this field of research could not be produced as soon as there is no clear distinction between the observer and the observed. Indeed, there is no way to know, for epistemological reasons, if what is observed is an effect induced by the experimenter (due to a possible psi influence) or a characteristic of the phenomena independently of the experimenter. The distinction between expectations and reality is then unclear, and the psi researcher using this approach can only become a modern version of Sisyphus[5] as it will be shown later.

TWO EXAMPLES FROM THE LITERATURE: THE GANZFELD PROTOCOL AND THE Bem STUDIES

The Ganzfeld experiment – the most classical protocol in psi research (Bem and Honorton, 1994) – provides an interesting example of the previous reasoning. During this experiment, a participant (the receiver) is comfortably seated in a chair, wearing a mask showing him a uniform visual field (usually red) while listening to a white noise. The participant is immersed in a constant and neutral sensory field that rapidly engenders an altered state of consciousness supposed to favor psi perceptions (Storm et al., 2010). After 20 min, the participant tries to find, among several images or videos (usually, the correct target and three alternate targets), the one that was "transmitted" telepathically by another participant (the sender) located in another room. As shown by the literature, an average success rate of 33% in the receiver's choice has been obtained instead of 25% by chance alone (Storm et al., 2010). A significant correlation between the choices of the receivers and the targets shown to the senders has thus been demonstrated. But what else has been learned and does this experiment demonstrate the existence of telepathy?

No, because it cannot be proven in a definite manner. Different competing interpretations could be proposed but they cannot be isolated or confirmed. There is indeed no way to propose a *falsifiable claim* or set-up a crucial experiment (Popper, 1934). For example, is it a telepathic effect (the target is "sent" by the sender and "received" by the receiver) or a precognitive effect (the receiver actually perceives the target from the "future," when presented with the four possible targets at the end of the session and asked to choose)? To test this hypothesis, an experiment can be set-up in which a feedback is given to half of the receivers (they see the correct target after the Ganzfeld state) and no feedback is given to half of the receivers (they do not see the correct target and, instead, a blind judge evaluates the correlation between the receiver's mentations and the four possible targets)[6]. If we suppose that a significant effect is obtained only for feedback trials, does it prove the precognitive hypothesis? Not so, because even if this experiment was replicated 50 times with the same results and using the best experimental conditions, there is no way to know if this effect is a consequence

[5]As it will be suggested later in this paper, there may be a possibility for psi researchers to avoid this path.

[6]Actually, 18% of Ganzfeld studies have been set-up without a sender and most of them still get significant results (Storm et al., 2010).

of (1) a precognitive effect, (2) a psi influence from the participants, (3) a psi effect from the experimenter on the computer that chooses the target, and (4) many other options! Because as soon as there is no clear *cut* between the experimenter, the participant, and the methodology, everything becomes possible.

Another possibility is that some experimenters are consistently able to influence the experimental data and so gain significant results. This variable itself has been the focus of considerable research (Broughton, 1979; Palmer, 1997; Parker and Miller, 2014)[7]. It can be tested during a Ganzfeld experiment by working with 10 different experimenters and by comparing their results. But even if a correlation between (for example) the experimenter's belief (a measure of its possible psi influence on the setup) and the Ganzfeld effect size is found, no clear conclusion could be reached. It could be (1) the effect of the experimenter, (2) the effect of the analyst of the study (the one who look at the data first), or (3) many other potential explanations! This is again the same problem: if there is no clear cut between the experimenter, the participants, and the scientific set-up, there is no falsifiable claim. Importantly, *this problem is infinite*: when new variables are introduced, without a clearly falsifiable hypothesis (due to the absence of an epistemological boundary between the observer and what is observed), the problem exists. This lack of clear attributable causation is relevant to the results of most psi experiments and that's *why the psi effects are probably not what they look like*. They can be considered as inconclusive from a scientific point of view. Thus, there is no way to be sure that the Ganzfeld is a transfer of information between two people or that precognition is the ability to extract information from the future. Psi studies actually just show that significant correlations between two variables – an intention and a measure – will emerge and take different forms depending on the conditions of the experimental setting.

The recent experiments reported by Bem (2011) on the anomalous anticipation of random events illustrate perfectly the problem we have just described. Bem (2011), publishing in a mainstream psychological journal, the *Journal of Personality and Social Psychology*, presented the results of four classical psychology experiments (e.g., priming task) that he has "reversed" in order to see if the participants would be influenced by stimuli not from the past but from the *future*. This publication induced considerable controversy between skeptics and proponents in the academic community and even in several mainstream media (Bem et al., 2011; Wagenmakers et al., 2011; Ritchie et al., 2012). We can sum up the different steps of this research paradigm: (1) Bem (2011) shows a seemingly reliable psi effect in controlled condition. It engenders critical reactions from the mainstream community (Wagenmakers et al., 2011). In a way, these critics are right: they probably "feel" that it does not look like a classical effect and maybe understand intuitively the potential consequences of this result for the whole field of psychology. (2) The experiment is then replicated in different settings (online, different populations, different stimuli, etc.)[8]. Sometimes it works, sometimes it does not work (Galak et al., 2012; Ritchie et al., 2012). (3) A meta-analysis is finally published with a significant, but smaller effect size (*es* = 0.14), than the original publication (*es* = 0.20; Bem et al., 2015)[9]. Is it because: (a) the effect actually does not exist; (b) the setting of the new experiments was less appropriate; (c) the replication process in itself decreases the effect? (Lucadou, 1995; Kennedy, 2003); or (d) the effect is actually the consequence of an experimenter effect described above? (Kennedy and Taddonio, 1976; Broughton, 1979; Palmer, 1997; Smith, 2003; Parker and Miller, 2014). A fourth step is thus necessary; and (4) a new study is concurrently conducted in different labs to study the impact of experimenter's individual differences on the results. But this effect could actually depend on the analyst (i.e., the one who look at the results for the first time or analyze the data) and not the experimenter (West and Fisk, 1953). So, the next step of this research paradigm might be (5) to test the magnitude of the effect with 10 different experimenters and 10 different analysts.

These different steps show the possible impediments which plague psi researchers in their efforts to prove the existence of psi, in addition to negative results and criticism from mainstream colleagues. And even if the researcher manages to get significant results at every step – as done by Bem (2011) for many years[10] –, there will always be new demands from the mainstream community: more control of the experimental setting, more experiments, more labs, more statistical tools, etc., (Wagenmakers et al., 2015)[11]. Even when proponents manage to agree a clear protocol with skeptics, and then obtain significant results, which has been the case with the Ganzfeld (Hyman and Honorton, 1986; Bem and Honorton, 1994), it is never enough. The underlying problem is that even if a significant effect is found at each step, there is no way to conclude anything about the nature of the effect and consequently no way to produce scientific knowledge about the *source* of psi (Broughton, 1979; Palmer, 1997): is it from the participants? From the experimenter? Is it from each experimenter separately? Or is it a stronger influence from the first one who analyze the data? Or, maybe that the one who has originally conceived the experiment? Are there degrees of influence between the

[7]Some researchers have already tried to do this with an experiment in which the participants had to mentally influence the electrodermal activity of another participant (Schlitz et al., 2006).

[8]For example, we have replicated the retro-priming experiment designed by Bem four times with a total of 344 participants (t = 0.92; d = 0.049; CI = −0.016, 0.16; p = 0.360; Rabeyron and Watt, 2010; Rabeyron, 2014; Rabeyron et al., 2018). One of the experiments has also been conducted with a population composed of artist in order to increase the effect but no significant result has been obtained (Rabeyron et al., 2018).

[9]This meta-analysis also shows that exactly replications are significant (*es* = 0.8) and that the effect mainly comes from fast thinking protocol (*es* = 0.11). The more successfully replicated protocol has been the precognitive detection of reinforcement (*es* = 0.14). For a critic of this meta-analysis, see Lakens (2015).

[10]Bem had already proposed significant results about psi since the 1990s with the Ganzfeld studies (Bem and Honorton, 1994), even after taking into account remarks from the critics.

[11]The French philosopher Méheust (1999) proposes a metaphor to describe this process close to the Sisyphus image. He describes psi research as a sweater constantly unravels. Each new generation has to "knit" again in order to maintain the existence of the sweater because new proofs are constantly necessary depending on the evolution of scientific methodology.

experimenter and the participants depending on the type of experiment? But also, is it an effect in the present (during the experiment) or is it an influence from the future (after the experiment), or even the past, if psi can transcend space and time? There are no definite answers to these questions, whatever results are obtained and unfortunately, there is, to our knowledge, no way to answer these questions because there are only *plausible interpretations*. Fundamentally, the problem is that the usual epistemological frame of research is not adequate when considering psi proprieties.

In this regard, it might be relevant to stop doing research whose aim is to prove the existence of psi using classical (scientific method) setting, because it does not really make sense from an epistemological point of view. It may be argued that this methodology cannot produce anything new even with large financing and the passage of a century of research. Nevertheless, these experiments are relevant in terms of *ritual*. A selection of classical psi experiments can be used – e.g., the Ganzfeld, dreams or presentiment studies – as illustrations of a phenomena, recognizing their "limitations" and without believing that something new explanation will be emerge from them. If they are conducted with enough intention, motivation, novelty, and creativity, these experiments should continue to produce significant results. Their interest is mainly strategic because it gives the opportunity to show that psi can be replicated and produce significant effects in controlled conditions as shown by several meta-analysis (Cardeña, 2018).

Nevertheless, these "demonstration studies" might even be more complicated if we also take into account that psi effects tend to disappear when the same experiment is replicated, which is described as the *elusive* nature of psi (Hansen, 2001; Kennedy, 2003). When a psi experiment is set-up, a distinction between two variables or a hypothesis is proposed (true/false). If another experiment uses the same hypothesis, many researchers have reported that the effect tends to disappear (Kennedy, 2003)[12]. In this regard, using the same hypothesis twice for a psi experiment could be like asking a comedian trying to make the public laugh with the same joke twice. Psi interactions seem to be the expression of a novelty and novelty, by definition, can be new only once. It might explain the strange results – inversion, displacement, and disappearance of the effect – that appear when the same experiment is replicated (Lucadou, 1995). In order to avoid this difficulty, and following the Sisyphus metaphor, some researchers take a small rock (an experiment), push it up the mountain, and then do this with another rock (another experiment), but they do not push the same rock twice to avoid too much "resistance" that would result from the replication of the same experiment. They do not do this because the effect does not exist, as suggested by Wiseman (2010), but as the consequence of the fact that this is the only way to maintain the effect.

Due to these different epistemological difficulties, new knowledge in the field of psi research based on "classical" protocol would not be reliable and even those related to the understanding of psi processes. For example, if personality traits are correlated with psi, how to be sure this is not the consequence of the psi influence from the experimenter about which personality traits he believes favor psi? This is the same for all parameters that could be correlated with psi results[13]. These researches cannot produce scientific knowledge and so may be considered as a waste of time and energy in the same way as Sisyphus spends all his time doing a useless and infinite task. Some researchers in the field have recognized this problem and have stopped doing this kind of experimental research (Eisenbud, 1966, 1983). Others understand this problem and try to find a way to avoid it with specific experimental set-up (Lucadou et al., 2007; Walach and von Stillfried, 2011). Others acknowledge this problem but continue to do experimental research like this because this approach is relevant from a "political" and strategic point of view (Radin, 2018). They know intuitively how to conduct experiments in order to keep "alive" a psi effect in spite of its profoundly elusive nature. They suppose that these experiments can be useful to convince the whole scientific community, and a larger audience, if they are conducted in a sufficient rigorous way. This could be considered as a pragmatic approach using the wrong tools to show something that might be true. And other researchers do not understand this problem and continue to do this kind of research, in the same manner as Rhine (1966) used to do, because they do not feel that there are other options. It would be like a woodsman trying to fell trees with a feather saying that he continues to do so because this is the only tool he has. If they are lucky, despite the inappropriate nature of the research methodology, they will occasionally obtain significant results, but will also obtain null results. If they are resilient, they will do this during all their careers and become Sisyphus, trying to convince a scientific community who do not believe in the existence of what they study. Not surprisingly, some of them will stop doing parapsychological research and even can become skeptic (Blackmore, 1987).

REPLICABILITY CRISIS, DECLINE EFFECT, AND PSI

As mentioned in the introduction, psychology and medicine have been confronted for more than 10 years by what has been called the replication or replicability crisis (Maxwell et al., 2015). It shows that the magnitude of the effect sizes in replications of psychology experiments is half the size of the original studies and that only 36% of the effect may be replicable

[12]Which could, of course, be interpreted more simply as the proof of the non-existence of psi and the "elusive argument" might be considered as a *post-hoc* hypothesis to explain non-significant results. This is why currently researchers try to demonstrate this elusive aspect of psi (Maier et al., 2018; Maier and Dechamps, 2018), but is it possible to demonstrate something that is supposed to be elusive?

[13]For example, a line of research has been developed to determine whether the nature of the targets during a psi experiment could influence the quality of the results. A significant correlation has been found between the descriptions of the participants and the degree of entropy of the target (e.g., a picture) from an informational point of view (May et al., 2000). But is it a "real" effect or the psi consequence of the belief of the experimenters (who as physicists interpret what is going on in terms familiar to them)?

(Nosek et al., 2015). The same problem is also true for other domains, especially medical and psychotherapy research (Ioannidis, 2016a,b). This is, of course, a huge problem that many researchers are trying to solve. One of the hypotheses to explain this situation is that these results could be the consequence of QRPs (John et al., 2012) and that most of these studies would not have been significant if they had been carried out with more rigor (Simmons et al., 2011)[14]. Consequently, the conduct of "Science" has to change by improving using – notably – pre-registration, better statistics, and the publication of null results.

The replication crisis has also underlined a phenomenon calls the "decline effect" (Schooler, 2011; Simmons et al., 2011; Protzko and Schooler, 2017). It shows that different effects in psychology and medicine tend to diminish with time and replication process (e.g., see Coyne and de Voogd, 2012; Gong and Jiao, 2019). In this regard, it is interesting to note that the psi community has reported such a decline effect a long time ago (Kennedy, 2003). Is it the same effect and what is its nature? Most researchers suppose that it is also the consequence of QRP[15] but a different hypothesis could be proposed; the underlying problem of this decline effect might be psi, if the latter exists. It means that a direct relationship between an intention and reality is possible. Consequently, when mainstream research is set-up, psi might come in the equation even if it is not invited to the party[16]. When researchers develop a new protocol or hypothesis, their expectations or intentions might, through psi, unconsciously induce a result which favors their view. Thus, when a new effect or a new treatment is tested – with, for example, a control group – the researcher might have a psi influence at various points in the research design[17], which could compromise the utility of the control group as a comparison condition.

As an example of how this unexpected influence could might affect data, May (1995) has developed the decision augmentation theory (DAT) to explain the results of studies in which participants had to influence, solely by intention, the output of a random number generator (Bösch et al., 2006). May supposes that there is no physical influence in this process and that the participant, using precognitive abilities, will choose (unconsciously) the right moment to push the button in order to get desired significant result[18]. In the same way, a researcher might unconsciously choose the right moment to start the study, choose the participants, collect the data, etc., in order to induce an effect in randomness. From this point of view, psi does not induce a transfer of energy or rely on a known physical force. It rather organizes reality in a discreet manner by ordering randomness. Consequently, some of the mainstream effects look like normal effects but they are not. It is only when other researchers – who may not have the same expectations, beliefs, or intentions – try to replicate them that these effects may mysteriously vanish. This would not be the effect of QRP, but the consequence of psi[19].

It could be argued that if this hypothesis is true, there is no possibility of accumulating *any* reliable scientific knowledge. But this is not the case because all observed effects are not attributable to psi; the latter acts as an "extended placebo" – that is, beyond the classical conception of placebo influence, see Lucadou, 2019 – that produce unexpected fluctuations in the data. But when a "real" and robust effect is replicated by different teams of researchers, it should resist if this is not a psi effect, and this might be what happened during the reproducibility project (Nosek et al., 2015). Thus, it is still possible to demonstrate the existence of "established," "real," or "classical" effects, laws, or forces (which probably concern the vast majority of reality) and, in this regard, the scientific model is still totally relevant. But it means that effect sizes around 0.10 and 0.20 – the usual magnitude reported in the psi literature – in experiments within many scientific domains, might actually be the consequence of psi[20].

The other interpretation of psi data – that psi does not exist, the "null hypothesis" (Alcock, 2003) – is also interesting from a psychological and sociological point of view. As proposed in introduction, it would suggest that hundreds of researchers (and notably more than 20 Nobel Prize winners) have been fooled for more than a century, even when using the most reliable tools of scientific research. The effect of these "illusory" results have been so convincing that they even led to practical applications (Schwartz, 2007; Mossbridge et al., 2014). For example, the United States government attempted to employ psi for more than 20 years during a program usually known as Stargate in which military personnel

[14]As explained by Cardeña (2019, p. 117–118), this problem of replicability because of statistical errors is not new and has already been described many times since the 1960s: "Just like tidal unveilings of flotsam, science discovers 'anew' errors of the third kind, Schlaiffer's term for the misuse and misinterpretation of statistical procedures besides the Type I and II errors (…) Have scientists become more insightful recently about the problems of mindless applications of statistical and research procedures? No, they are just reiterating ideas that have been around for more than a half century".

[15]The QRP hypothesis has been evaluated by Bierman (2016) employing the Ganzfeld database. He concludes that the results are probably inflated because of QRP but that they are still significant (p = 0.003). So the QRP hypothesis is not sufficient to explain the Ganzfeld results and, by implication, the results of many psi experiments.

[16]Classical mainstream experiments have been re-analyzed taking into account the psi hypothesis, leading sometimes to significant results. For example, Bierman (2000) has found a presentiment effect in data from classical Antonio Damasio experiments.

[17]The researchers have usually a particularly strong intention toward their results because their ability to publish in a good journal, and their own career, depends on the results they will get. Paradoxically, a scientific experiment could be considered as one of the best set-up to induce psi effect.

[18]For example, if the participant has to push a button in order to put a light in green or in red (the color of the light depends on the result of the RNG), he does not have a direct and mental influence on the RNG; he has actually choose the right time to push the button in order to select a random binary sequence (more 0 or more 1) associated with the right color.

[19]It could also be hypothesized that other controversial effects like homeopathy and different forms of complementary and alternative medicine practices (Hyland, 2003) could be the consequence of psi (Walach, 2000; Lucadou, 2019).

[20]If this hypothesis is true, it is not only important from a scientific point of view, but it also means that large amount of money invested in scientific research are lost because they actually concern psi effects. In this regard, it would be fundamental to find a way to discriminate "classical" from "psi" effects in order to avoid this problem, which can have serious consequences, especially in medical research.

were selected on basis of their supposed psi-abilities to acquire information (e.g., about Russian military sites) at a distance (in space and time; Hyman, 1996; Utts, 1996; May et al., 2018)[21]. The null hypothesis would mean that staff of the best United States intelligence agencies (CIA, NSA, etc.), a number of military officers working on this program (some of whom were decorated with the legion of merit; McMoneagle and May, 2014), top scientists who have examined the project (notably a past president of the *American Statistical Association*; Utts, 1996), and even the president of the United States (Jimmy Carter admitted that a lost military plane Tu-22 has been found thanks to the Stargate program) have been fooled by the results of 504 military operations over almost 20 years (1973–1995). If this interpretation of significant results in psi experiments is accepted, it may follow that other areas of "reputable" research, involving many researchers, could also produce illusory results.

TO CONCLUDE: PERSPECTIVES FOR FUTURE RESEARCH

Psi studies are particularly interesting because whatever the reaction to the question "does psi exist?" (Bem and Honorton, 1994), their results affect the whole of psychology. If *psi does not exist*, significant results for nearly a century have only been obtained by methodological errors, self-deception, fraud, and questionable research practices. How could we avoid such a problem? Since the beginning of the replicability crisis, several solutions have been proposed – pre-registration of study designs, Bayesian statistics, larger *N*, funnel plots, *p*-curve analysis, prospective meta-analysis, adversial collaborations, etc., (Bateman et al., 2005) – which could show, at the end, non-significant results in the field of psi studies, revealing that psi was only an illusion. A pre-registration registry has already been set-up in the field of psi research[22] (Watt and Kennedy, 2015, 2017, 2019) as well as statistical guidelines for empirical studies (Tressoldi and Utts, 2015; Kennedy, 2016; Utts and Tressoldi, 2019). Pre-submission to scientific journals which accept a paper on methodological grounds prior to results should also be promoted. In this regard, a "transparent psi project" is currently being conducted which follows these recommendations[23]. Such an approach might be extended to other psi paradigms to confirm or deny the significant results of several meta-analysis (Sherwood and Roe, 2003; Storm et al., 2010; Mossbridge et al., 2012).

On the contrary, if *psi does exist*, it means that human consciousness can interact with its environment beyond the usual boundaries of space and time, which has fundamental consequences for the way research is conducted in psychology, including psi research (as demonstrated by the psi paradox). As already mentioned, the results of experimental psi research have shown, since their beginning, strange patterns in the data (displacement, reversal, etc.) called notably psi-missing (Rhine, 1952) and elusiveness (Kennedy, 2003). A solution might be to consider these patterns not as an obstacle – or just the effect of randomness (Wiseman, 2010) – but rather as a way to better understand psi and its properties[24]. Following this idea, an original line of research has been initiated by the physicist and psychologist Walter Von Lucadou with the "Model of Pragmatic Information" (MPI; Lucadou, 1995; Lucadou et al., 2007). In this model, psi is considered as being something profoundly different to known macro-physical effects and causation, not relying on transfer of information but rather a form of entanglement process depending on the underlying nature of reality (Atmanspacher and Fuchs, 2017; Atmanspacher and Fach, 2019)[25]. A brief metaphor might be useful here. A psi experiment is like an egg where the shell forms an enclosed organizational system. It may be possible to maintain a psi effect as long as the organizational closure is not broken, that is as long as the egg is not broken to see what is inside. In this interpretation, the psi interactions are possible as long as the observer does not interfere with the system (Houtkooper, 2002). Once the system is observed, "the game is over." This would explain why the source of psi cannot be determined precisely because the determination process would destroy the necessary conditions for the emergence of psi. It also underlines the importance of uncertainty associated with the source of psi. When the latter is used for a transfer of information, the psi effect would be suppressed, especially when attempts are made to replicate

[21]Such approach has also been used for searching missing persons (Schwartz, 2007), archeology (Schwartz, 2001), and financial investing (Bierman and Rabeyron, 2012).

[22]See https://koestlerunit.wordpress.com/study-registry/

[23]The results will be published in *Royal society open science*. More details here: https://psyarxiv.com/uwk7y/

[24]Another original approach consists in determining biological and genetic markers of psi. The aim is then not to demonstrate the existence of psi but rather to describe its biological foundations. For example, phenomenological and neurobiological aspects of synesthesia can be evaluated (Eagleman et al., 2007), and psi might have a lot in common with synesthesia (Simmonds-Moore et al., 2019). So, if the profound nature of psi cannot be explained, reliable markers correlated with its expression using fMRI or EEG studies could be determined in order to find neuro-correlates of psi (Moulton and Kosslyn, 2008). But the level of research in this domain has to be improved (Acunzo et al., 2013), and it should be proven that these correlations in fMRI results are not psi effects. Genetic analysis could also be relevant to detect selected participants and represent a particularly promising area of research. It is also interesting to note that a number of important discoveries have been made by working with simple organisms. A famous example has been given by a Nobel Prize winner, Kandel (2007), who has discovered molecular aspects of memory processes thanks to *Aplysia californica* – a small sea slug – equipped with a very simple nervous system (only 20,000 neurons). If psi is not a specificity of human consciousness and is shared with other species (Sheldrake, 2004), it could be relevant to look for the equivalent of *A. californica* for psi research, that is a simple biological system (plants, animals, etc.) that would allow to manipulate different variables and go beyond correlational studies.

[25]In this regard, it seems that what is observed at a macro level is close to what is described by physicists at a quantum level about the influence of the observer. The MPI relies on a mathematical formalism, the "Generalized (or Weak) Quantum Theory" (GQT; Filk and Römer, 2011; Walach and von Stillfried, 2011) which use quantum mathematical model as a metaphor. But this model does not pretend that psi is a quantum phenomenon. It rather supposes that several aspects of quantum formalism (e.g., non commutativity), which are very useful in physic, might also be relevant to describe psi processes.

exactly the same experimental set. This is what Lucadou calls the "Non-Transmission Axiom" (Lucadou et al., 2007).

Consequently, Lucadou has tried to set-up an experiment in which this type of effect might be maintained by keeping a sufficient level of uncertainty in the system. This experiment uses the "Correlation Matrix Method" (CMM) in which the global number of correlations between the participants and an experimental task (associated with a RNG) is predicted, but not the location of such correlations in the correlational matrix (Lucadou, 2015; Flores et al., 2018; Walach et al., 2019). The non-transmission axiom could also explain the decline effect and the oscillating trends in the data (Pallikari and Boller, 1997; Maier et al., 2018; Maier and Dechamps, 2018). This last aspect is particularly interesting because these oscillating patterns might be detected, demonstrated, and analyzed when they are compared with classical effects (Rabeyron, 2014).

This line of research appears as an interesting example of what could be conceived as an example of "postmodern psychology" which takes into account the complexity of human consciousness, and more precisely postulates a potential entanglement between the observer and what is observed. It also shows how psi might be implicated in the "hard problem" of consciousness (Chalmers, 2007) or the "problem of measurement" (Wigner, 1963). Even if the possibility that psi exists sounds very implausible to many (Wiseman, 2010; Reber and Alcock, 2020), and as proposed recently by Schooler et al. (2018), a neutral and respectful approach to this topic might open heuristic debates within the wider field of psychology concerning the replicability crisis and the nature of consciousness.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Acunzo, D. J., Evrard, R., and Rabeyron, T. (2013). Anomalous experiences, psi and functional neuroimaging. *Front. Hum. Neurosci.* 7:893. doi: 10.3389/fnhum.2013.00893

Alcock, J. E. (2003). Give the null hypothesis a chance: reasons to remain doubtful about the existence of psi. *J. Conscious. Stud.* 10, 29–50.

Alcock, J. E., Burns, J., and Freeman, A. (2003). *Psi wars: Getting to grips with the paranormal*. Charlottesville: Imprint Academic.

Atmanspacher, H., and Fach, W. (2019). Exceptional experiences of stable and unstable mental states, understood from a dual-aspect point of view. *Philosophies* 4:7. doi: 10.3390/philosophies4010007

Atmanspacher, H., and Fuchs, C. A. (2017). *The Pauli-Jung conjecture and its impact today*. London: Andrews UK Limited.

Bateman, I., Kahneman, D., Munro, A., Starmer, C., and Sugden, R. (2005). Testing competing models of loss aversion: an adversarial collaboration. *J. Public Econ.* 89, 1561–1580. doi: 10.1016/j.jpubeco.2004.06.013

Bem, D. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *J. Pers. Soc. Psychol.* 100, 407–425. doi: 10.1037/a0021524

Bem, D., and Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychol. Bull.* 115, 4–18. doi: 10.1037/0033-2909.115.1.4

Bem, D., Tressoldi, P., Rabeyron, T., and Duggan, M. (2015). Feeling the future: a meta-analysis of 90 experiments on the anomalous anticipation of random future events. *F1000Res* 4:1188. doi: 10.12688/f1000research.7177.2

Bem, D., Utts, J., and Johnson, W. O. (2011). Must psychologists change the way they analyze their data? *J. Pers. Soc. Psychol.* 101, 716–719. doi: 10.1037/a0024777

Bierman, D. J. (2000). Anomalous baseline effects in mainstream emotion research using psychophysiological variables. *J. Parapsychol.* 64:239.

Bierman, D. J., and Rabeyron, T. (2012). "Can psi research sponsor itself? Simulations and results of an automated ARV-casino experiment." in *Paper presented at the 56st annual convention of the parapsychological association*.

Bierman, D. J., Spottiswoode, J. P., and Bijl, A. (2016). Testing for questionable research practices in a meta-analysis: an example from experimental parapsychology. *PLoS One* 11:e0153049. doi: 10.1371/journal.pone.0153049

Blackmore, S. (1987). The elusive open mind: ten years of negative research in parapsychology. *Skeptical Inquirer* 11, 244–255.

Bösch, H., Steinkamp, F., and Boller, E. (2006). Examining psychokinesis: the interaction of human intention with random number generators–a meta-analysis. *Psychol. Bull.* 132, 497–523. doi: 10.1037/0033-2909.132.4.497

Braude, S. E. (2007). *The gold leaf lady and other parapsychological investigations*. Chicago: University of Chicago Press.

Broughton, R. S. (1979). Repeatability and experimenter effect: are subjects really necessary? *Parapsychol. Rev.* 10, 11–14.

Cardeña, E. (2018). The experimental evidence for parapsychological phenomena: a review. *Am. Psychol.* 73, 663–677. doi: 10.1037/amp0000236

Cardeña, E. (2019). Errors of the third kind. *J. Parapsychol.* 83, 117–120.

Cardeña, E., Palmer, J., and Marcusson-Clavertz, D. (2015). *Parapsychology: A handbook for the 21st century*. Jefferson: McFarland.

Carpenter, J. C. (2012). *First sight: ESP and parapsychology in everyday life*. Lanham: Rowman & Littlefield Publishers.

Chalmers, D. (2007). "The hard problem of consciousness" in *The Blackwell companion to consciousness*. eds. S. Schneider and M. Velmans (Chichester: Wiley-Blackwell).

Chalmers, A. F. (1979). *What is this thing called science?: An assessment of the nature and status of science and its methods*. St. Lucia: University of Queensland Press.

Coyne, J. C., and de Voogd, J. N. (2012). Are we witnessing the decline effect in the Type D personality literature? What can be learned? *J. Psychosom. Res.* 73, 401–407. doi: 10.1016/j.jpsychores.2012.09.016

Eagleman, D. M., Kagan, A. D., Nelson, S. S., Sagaram, D., and Sarma, A. K. (2007). A standardized test battery for the study of synesthesia. *J. Neurosci. Methods* 159, 139–145. doi: 10.1016/j.jneumeth.2006.07.012

Eisenbud, J. (1966). *The world of Ted serios: "Thoughtographic" studies of an extraordinary mind*. New York: William Morrow.

Eisenbud, J. (1983). *Parapsychology and the unconscious*. New York: North Atlantic Book.

Feyerabend, P. (1975). *Contre la méthode*. Paris: Le Seuil.

Filk, T., and Römer, H. (2011). Generalized quantum theory: overview and latest developments. *Axiomathes* 21, 211–230. doi: 10.1007/s10516-010-9136-6

Flores, A. B., Tierney, I., and Watt, C. A. (2018). Where mind connects with matter: replicating the correlation matrix method. *J. Parapsychol.* 82, 107–108.

Galak, J., LeBoeuf, R. A., Nelson, L. D., and Simmons, J. P. (2012). Correcting the past: failures to replicate psi. *J. Pers. Soc. Psychol.* 103, 933–948. doi: 10.1037/a0029709

Gong, Z., and Jiao, X. (2019). Are effect sizes in emotional intelligence field declining? A meta-meta analysis. *Front. Psychol.* 10:1655. doi: 10.3389/fpsyg.2019.01655

Hansel, C. E. M. (1961). A critical analysis of the Pearce-Pratt experiment. *J. Parapsychol.* 25, 87–91.

Hansen, G. (2001). *The trickster and the paranormal*. Philadelphia: Xilibiris Corporation.

Holt, N. (2012). "The muse in the machine: creativity, anomalous experiences and mental health" in *Exceptional experience and health*. ed. C. Simmonds-Moore (Jefferson: McFarland).

Houtkooper, J. M. (2002). Arguing for an observational theory of paranormal phenomena. *J. Sci. Explor.* 16, 171–185.

Hyland, M. E. (2003). Extended network generalized entanglement theory: therapeutic mechanisms, empirical predictions, and investigations. *J. Altern. Complement. Med.* 9, 919–936. doi: 10.1089/107555303771952262

Hyman, R. (1996). Evaluation of a program on anomalous mental phenomena. *J. Sci. Explor.* 10, 321–352.

Hyman, R. (2010). "Parapsychology's achilles heel: persistent inconsistency" in *Debating psychic experience*. eds. S. Krippner and H. L. Friedman (Santa Barbara: Praeger).

Hyman, R., and Honorton, C. (1986). A joint communiqué: the psi ganzfeld controversy. *J. Parapsychol.* 50, 351–364.

Ioannidis, J. P. (2016a). Most psychotherapies do not really work, but those that might work should be assessed in biased studies. *Epidemiol. Psychiatr. Sci.* 25, 436–438. doi: 10.1017/S2045796015000888

Ioannidis, J. P. (2016b). Why most clinical research is not useful. *PLoS Med.* 13:e1002049. doi: 10.1371/journal.pmed.1002049

Irwin, H. J., and Watt, C. (2007). *An introduction to parapsychology*. Jefferson: McFarland.

John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* 23, 524–532. doi: 10.1177/0956797611430953

Kandel, E. R. (2007). *In search of memory: The emergence of a new science of mind*. New York: Norton & Company.

Kennedy, J. E. (2003). The capricious, actively evasive, unsustainable nature of psi: a summary and hypotheses. *J. Parapsychol.* 67, 53–74.

Kennedy, J. E. (2016). Is the methodological revolution in psychology over or just beginning. *J. Parapsychol.* 80, 156–168.

Kennedy, J. E., and Taddonio, J. L. (1976). Experimenter effects in parapsychological research. *J. Parapsychol.* 40, 1–33.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago press.

Lakens, D. (2015). *Why a meta-analysis of 90 precognition studies does not provide convincing evidence of a true effect*. The Winnower, 6.

Lawrence, T. R. (1993). "Gathering in the sheep and goats: a meta-analysis of forced-choice sheep-goat ESP studies, 1947–1993." in *Paper presented at the 36st Annual Convention of the Parapsychological Association*; 1993, Parapsychological Association.

Lucadou, W. (1995). The model of pragmatic information (MPI). *Eur. J. Parapsychol.* 11, 58–75.

Lucadou, W. (2015). "The correlation matrix method (CMM): A new light upon the repeatability problem in parapsychology." in *Paper for the 58th Annual Convention of the Parapsychological Association and the 39th SPR International Conference*; 2015, Parapsychological Association.

Lucadou, W. (2019). Homeopathy and the action of meaning: a theoretical approach. *J. Sci. Explor.* 33, 213–254. doi: 10.31275/2019.1343

Lucadou, W., Romer, H., and Walach, H. (2007). Synchronistic phenomena as entanglement correlations in generalized quantum theory. *J. Conscious. Stud.* 14, 50–74.

Maier, M. A., and Dechamps, M. C. (2018). Observer effects on quantum randomness: testing micro-psychokinetic effects of smokers on addiction-related stimuli. *J. Sci. Explor.* 32, 265–297. doi: 10.31275/2018.1250

Maier, M. A., Dechamps, M. C., and Pflitsch, M. (2018). Intentional observer effects on quantum randomness: a Bayesian analysis reveals evidence against micro-psychokinesis. *Front. Psychol.* 9:379. doi: 10.3389/fpsyg.2018.00379

Maxwell, S. E., Lau, M. Y., and Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *Am. Psychol.* 70, 487–498. doi: 10.1037/a0039400

May, E. C., Marwaha, S. B., Broughton, F., Cohen, W. S., and Broughton, R. S. (2018). *The star gate archives: Reports of the United States government sponsored psi program, 1972–1995: Volume 1: Remote viewing, 1972–1984*. Jefferson: McFarland & Company.

May, E. C., Spottiswoode, S., and Faith, L. (2000). The correlation of the gradient of Shannon entropy and anomalous cognition: toward an AC sensory system. *J. Sci. Explor.* 14, 53–72.

May, E. C., Utts, J. M., and Spottiswoode, S. J. P. (1995). Decision augmentation theory: applications to the random number generator database. *J. Sci. Explor.* 9, 453–488.

McMoneagle, J. W., and May, E. C. (2014). "The possible role of intention, attention and expectation in remote viewing" in *Anomalous cognition: Remote viewing research and theory*. eds. E. C. May and S. B. Marwaha (Jefferson: McFarland), 368–376.

Méheust, B. (1999). *Somnambulisme et mediumnité*. Paris: Les Empêcheurs de penser en rond.

Méheust, B. (2003). *Un voyant prodigieux: Alexis Didier (1826–1866)*. Paris: Les Empêcheurs de penser en rond.

Mossbridge, J. A., and Radin, D. (2018). Precognition as a form of prospection: a review of the evidence. *Psychol. Conscious.* 5, 78–93. doi: 10.1037/cns0000121

Mossbridge, J., Tressoldi, P., and Utts, J. (2012). Predictive physiological anticipation preceding seemingly unpredictable stimuli: a meta-analysis. *Front. Psychol.* 3:390. doi: 10.3389/fpsyg.2012.00390

Mossbridge, J., Tressoldi, P. E., Utts, J., Ives, J., Radin, D., and Jonas, W. (2014). Predicting the unpredictable: critical analysis and practical implications of predictive anticipatory activity. *Front. Hum. Neurosci.* 8:146. doi: 10.3389/fnhum.2014.00146

Moulton, S., and Kosslyn, S. M. (2008). Using neuroimaging to resolve the psi debate. *J. Cogn. Neurosci.* 20, 182–192. doi: 10.1162/jocn.2008.20009

Nosek, B. A., Aarts, A. A., Anderson, C. J., Anderson, J. E., Kappes, H. B., and Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716

Pallikari, F., and Boller, E. (1997). Further evidence for a statistical balancing in probabilistic systems influenced by the anomalous effect of conscious intention. *J. Soc. Psychol. Res.* 62, 114–137.

Palmer, J. (1997). The challenge of experimenter psi. *Eur. J. Parapsychol.* 13, 110–125.

Palmer, J. (2016). Hansel's ghost: resurrection of the experimenter fraud hypothesis in parapsychology. *J. Parapsychol.* 80, 5–16.

Parker, A., and Millar, B. (2014). Revealing psi secrets: successful experimenters seem to succeed by using their own psi. *J. Parapsychol.* 78, 39–55.

Pashler, H., and Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspect. Psychol. Sci.* 7, 531–536. doi: 10.1177/1745691612463401

Popper, K. (1934). *The logic of scientific discovery*. New York: Routledge.

Protzko, J., and Schooler, J. W. (2017). "Decline effects: types, mechanisms, and personal reflections" in *Psychological science under scrutiny: Recent challenges and proposed solutions*. eds. S. O. Lilienfeld and I. D. Waldman (New Yersey: Wiley-Blackwell).

Rabeyron, T. (2014). Retro-priming, priming, and double testing: psi and replication in a test-retest design. *Front. Hum. Neurosci.* 8:154. doi: 10.3389/fnhum.2014.00154

Rabeyron, T. (2020). *Clinique des expériences exceptionnelles*. Paris: Dunod.

Rabeyron, T., and Loose, T. (2015). Anomalous experiences, trauma, and symbolization processes at the Frontiers between psychoanalysis and cognitive neurosciences. *Front. Psychol.* 6:1926. doi: 10.3389/fpsyg.2015.01926

Rabeyron, T., Rowe, C., Mousseau, M. C., and Deledalle, A. (2018). Anomalous experiences, mental health, and creativity: is psi the missing link? *J. Conscious. Stud.* 25, 207–232.

Rabeyron, T., and Watt, C. (2010). Paranormal experiences, mental health and mental boundaries, and psi. *Pers. Individ. Differ.* 48, 487–492. doi: 10.1016/j.paid.2009.11.029

Radin, D. (2006). *Entangled minds: Extrasensory experiences in a quantum reality*. New York: Paraview Pocket Books.

Radin, D. (2018). *Real magic: Ancient wisdom, modern science, and a guide to the secret power of the universe*. New York: Harmony.

Rao, K. R., and Palmer, J. (1987). The anomaly called psi: recent research and criticism. *Behav. Brain Sci.* 10, 539–551. doi: 10.1017/S0140525X00054455

Reber, A. S., and Alcock, J. E. (2020). Searching for the impossible: parapsychology's elusive quest. *Am. Psychol.* 75, 391–399. doi: 10.1037/amp0000486

Rhine, J. B. (1952). The problem of psi-missing. *J. Parapsychol.* 16, 90–129.

Rhine, J. B., Pratt, J. G., Smith, B. M., Stuart, C. E., and Greenwood, J. A. (1966). *Extra-sensory perception after sixty years*. Boston: Bruce Humphries Publishers.

Rhine-Feather, S., and Schmicker, M. (2005). *The gift: The extraordinary experiences of ordinary people*. New York: St. Martin's Paperbacks.

Ritchie, S. J., Wiseman, R., and French, C. C. (2012). Failing the future: three unsuccessful attempts to replicate bem's 'retroactive facilitation of recall' effect. *PLoS One* 7:e33423. doi: 10.1371/journal.pone.0033423

Romero, F. (2019). Philosophy of science and the replicability crisis. *Philos Compass* 14:e12633. doi: 10.1111/phc3.12633

Savalei, V., and Dunn, E. (2015). Is the call to abandon p-values the red herring of the replicability crisis? *Front. Psychol.* 6:245. doi: 10.3389/fpsyg.2015.00245

Schlitz, M. J., and Honorton, C. (1992). Ganzfeld psi performance within an artistically gifted population. *J. Am. Soc. Psychol. Res.* 86, 83–98.

Schlitz, M. J., Wiseman, R., Watt, C., and Radin, D. (2006). Of two minds: skeptic-proponent collaboration within parapsychology. *Br. J. Psychol.* 97, 313–322. doi: 10.1348/000712605X80704

Schooler, J. (2011). Unpublished results hide the decline effect. *Nature* 470:437. doi: 10.1038/470437a

Schooler, J. W., Baumgart, S., and Franklin, M. (2018). Entertaining without endorsing: the case for the scientific investigation of anomalous cognition. *Psychol. Conscious.* 5, 63–77. doi: 10.1037/cns0000151

Schwartz, S. (2001). *The Alexandria project*. Bloomington: IUniverse.

Schwartz, S. (2007). *Opening to the infinite*. Virginia Beach: Nemoseen Media.

Sheldrake, R. (2004). "Animal psi" in *Parapsychology: Research on exceptional experiences.* ed. J. Henry (London: Routledge).

Sherwood, S. J., and Roe, C. A. (2003). A review of dream ESP studies conducted since the Maimonides dream ESP programme. *J. Conscious. Stud.* 10, 85–109.

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632

Simmonds-Moore, C. A., Alvarado, C. S., and Zingrone, N. L. (2019). A survey exploring synesthetic experiences: exceptional experiences, schizotypy, and psychological well-being. *Psychol. Conscious.* 6, 99–121. doi: 10.1037/cns0000165

Smith, M. (2003). The role of the experimenter in parapsychological research. *J. Conscious. Stud.* 10, 69–84.

Storm, L., Tressoldi, P., and Risio, L. D. (2010). Meta-analysis of free-response studies, 1992–2008: assessing the noise reduction model in parapsychology. *Psychol. Bull.* 136, 471–485. doi: 10.1037/a0019457

Thalbourne, M. A. (2000). Transliminality: a review. *Int. J. Parapsychol.* 11, 1–34.

Tressoldi, P. E., and Utts, J. (2015). "Statistical guidelines for empirical studies" in *Parapsychology: A handbook for the 21st century.* eds. E. Cardeña, J. Palmer and D. Marcusson-Clavertz (Jefferson: McFarland).

Utts, J. (1996). An assessment of the evidence for psychic functioning. *J. Sci. Explor.* 10, 3–30. doi: 10.30891/jopar.2018.03.10

Utts, J., and Tressoldi, P. E. (2019). Invited editorial methodological and statistical recommendations: option or necessity? *J. Parapsychol.* 83, 120–122. doi: 10.30891/jopar.2019.02.02

Wagenmakers, E. -J., Wetzels, R., Borsboom, D., Kievit, R. A., and van der Maas, H. L. (2015). "A skeptical eye on psi" in *Extrasensory perception: Support, skepticism, and science.* eds. E. C. May and S. B. Marwaha (Westport: Praeger).

Wagenmakers, E., Wetzels, R., Borsboom, D., and Van der Maas, H. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *J. Pers. Soc. Psychol.* 100, 426–432. doi: 10.1037/a0022790

Walach, H. (2000). Magic of signs: a non-local interpretation of homeopathy. *Br. Homeopath. J.* 89, 127–140. doi: 10.1054/homp.1999.0413

Walach, H., Horan, M., Hinterberger, T., and von Lucadou, W. (2019). Evidence for anomalistic correlations between human behavior and a random event generator: result of an independent replication of a micro-PK experiment. *Psychol. Conscious.* 7, 173–188. doi: 10.1037/cns0000199

Walach, H., and von Stillfried, N. (2011). Generalised quantum theory—basic idea and general intuition: a background story and overview. *Axiomathes* 21, 185–209. doi: 10.1007/s10516-010-9145-5

Watt, C., and Kennedy, J. E. (2015). Lessons from the first two years of operating a study registry. *Front. Psychol.* 6:173. doi: 10.3389/fpsyg.2015.00173

Watt, C., and Kennedy, J. E. (2017). Options for prospective meta-analysis and introduction of registration-based prospective meta-analysis. *Front. Psychol.* 7:2030. doi: 10.3389/fpsyg.2016.02030

Watt, C., and Kennedy, J. E. (2019). The American statistician special issue on statistical inference. *J. Parapsychol.* 83, 283–285. doi: 10.30891/jopar.2019.02.14

West, D. J., and Fisk, G. W. (1953). A dual ESP experiment with clock cards. *J. Soc. Psychol. Res.* 37, 185–197.

Wiggins, B. J., and Chrisopherson, C. D. (2019). The replication crisis in psychology: an overview for theoretical and philosophical psychology. *J. Theor. Philos. Psychol.* 39, 202–217. doi: 10.1037/teo0000137

Wigner, E. P. (1963). The problem of measurement. *Am. J. Phys.* 31, 6–15. doi: 10.1119/1.1969254

Wiseman, R. (2010). Heads i win, tails you lose: how parapsychologists nullify null results. *Skeptical Inquirer* 34, 36–39.

Witte, E. H., and Zenker, F. (2017). From discovery to justification: outline of an ideal research program in empirical psychology. *Front. Psychol.* 8:1847. doi: 10.3389/fpsyg.2017.01847

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 01 December 2020
doi: 10.3389/fpsyg.2020.597706

Prediction and Explanation in a Postmodern World

*Joachim I. Krueger**

Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, RI, United States

The experimental research paradigm lies at the core of empirical psychology. New data analytical and computational tools continually enrich its methodological arsenal, while the paradigm's mission remains the testing of theoretical predictions and causal explanations. Predictions regarding experimental results necessarily point to the future. Once the data are collected, the causal inferences refer to a hypothesis now lying in the past. The experimental paradigm is not designed to permit strong inferences about particular incidents that occurred before predictions were made. In contrast, historical research and scholarship in other humanities focus on this backward direction of inference. The disconnect between forward-looking experimental psychology and backward-looking historical (i.e., narrative) psychology is a challenge in the postmodern era, which can be addressed. To illustrate this possibility, I discuss three historical case studies in light of theory and research in contemporary psychology.

Keywords: prediction, explanation, experimentation, causal analysis, postmodernism

OPEN ACCESS

Edited by:
Barbara Hanfstingl,
University of Klagenfurt, Austria

Reviewed by:
Peter Holtz,
Leibniz-Institut für Wissensmedien (IWM), Germany
Joachim Funke,
Heidelberg University, Germany

***Correspondence:**
Joachim I. Krueger
joachim@brown.edu

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *21 August 2020*
Accepted: *22 October 2020*
Published: *01 December 2020*

Citation:
Krueger JI (2020) Prediction and Explanation in a Postmodern World. Front. Psychol. 11:597706. doi: 10.3389/fpsyg.2020.597706

INTRODUCTION

"Wo der Brotgelehrte trennt, vereinigt der philosophische Geist."
[Where the ordinary scholar divides, the philosophical spirit unites.]
– Friedrich Schiller, May 26, 1789, in his first lecture as chair of history at the University of Jena

Friedrich Schiller did not shy away from lofty goals. Envisioning a universal theory of history, he surely realized that his listeners were guided by more modest aspirations. Yet, lofty goals are useful, and often needed, because they appeal to our better ideals, even if the odds of reaching them are long. Many scientists and scholars today might agree that everything is ultimately connected. The *Kosmos*, as von Humboldt (1845–1858) taught, is one. In practice, however, it is difficult to work from a universal point of view, and the academy has found it expedient to create distinctive administrative units where researchers can ask narrow questions they find tractable given the theories and methods available in their fields. Administrators periodically affirm the need for interdisciplinary or translational research, and occasionally they provide funds to support it. Such initiatives are useful as they guard against a descent into a world in which members of different academic tribes no longer understand one another.

The descent into tribalism may be more threat than reality, but the differences in methods and modes of thinking among disciplines are striking. The challenge remains to find answers to the question of what different disciplines can offer one another and whether these gifts can be used to good effect. The research topic presented by *Frontiers in Psychology*, to which this article seeks to make a contribution, asks about "modern" and "postmodern" approaches, and how their differences might be overcome. Taking the terms "modern" and "postmodern" as they are commonly understood, the prospect of a full reconciliation seems remote (Bereiter, 1994). Postmodernism, as it emerged from French *théorie* sees itself as a revolution, with its *raison d'être* being the rejection of modernism (Pluckrose and Lindsay, 2020). Modernism, and

the Enlightenment from which is sprang, is a necessary condition for postmodernism. "If the enlightenment did not exist, postmodernists would have had to invent it" (Schmidt, 2002, p. 432). The mutual distrust runs deep. Much like the surrealists did not wish to compromise with the realists, or the Copernicans did not wish to "split the difference" with the Ptolemaians,[1] so postmodernists appear to have no use for a middle ground between their own views and those of the modernists. A key demand of postmodernism is the co-existence of multiple epistemologies and the rejection of any kind of value-based ranking (Feyerabend, 1975). As this demand amounts to a rejection of modernist "science" as a privileged paradigm (Adams St. Pierre, 2002), it is hard to see what a compromise might look like. Modernists for their part maintain that science is not just another epistemology, on a par, as Feyerabend would have it, with astrology or voodoo. Science does not reduce to scientism, as Hutchinson (2011) claimed it does. Science provides the tools to study the validity of astrology, voodoo, and intercessory prayer, whereas the inverse effort cannot get off the ground. The most penetrating criticisms of scientific methods come from within the scientific community itself, and they help to improve the enterprise.

PREVIEW

In this article, I accept the general legitimacy of the modernist conception of science while exploring questions touching on postmodern sensibilities. Specifically, I pose two questions that I consider to be tractable with standard epistemological tools. The first question, which is theoretical, is how the concepts of prediction and explanation as construed in the conventional hypothesis-testing experimental paradigm relate to the concept of explanation as seen in historical accounts of events that happened in the past and outside the laboratory, or "in the wild." Taking a Bayesian perspective (Mandel, 2014), I frame this issue as a question of "reverse inference" (Krueger, 2017b). After reviewing the theoretical tools, I present a stylized thought experiment to illustrate the probabilistic association between forward and reverse inferences.

The second question, which is practical in nature, is whether theory and evidence-based psychological reconstructions of individuals' past decisions or behaviors can enrich historical scholarship. To explore the potential contributions of psychological reconstruction, I present three case studies. The first case features Philipp von Hutten, a historical person; the second involves Gonzalo Guerrero, a man suspended between history and legend; the third involves Robinson Crusoe, a figure familiar from literature. All three men found themselves in challenging circumstances demanding life-and-death decisions. With regard to Hutten and Gonzalo, I ask whether experimental psychology can help explain some of their critical choices. For the case of Crusoe, I introduce theoretical tools provided by a heterodox branch of game theory in order to reconstruct the interpersonal power dynamics between him and Friday. There is a limited tradition of historical case studies guided by theory and evidence. Simonton (1998), for example, pioneered and refined historiometric analyses of individuals of great creative or productive achievement. Dörner and Güss (2011) described, analyzed, and evaluated Hitler's rigid pattern of strategic decision-making in light of psychological theories of cognition and personality. These efforts fall within the purview of "differential psychology," yielding inferences that are only loosely tied to theories of psychological processes. The present article seeks to explore possible contributions of experimental psychology.

To conclude this article, I revisit three conceptual issues, which, if resolved, can shed light on the linkages between prediction and explanation, and, by extension, the linkages between modern and postmodern psychology. I ask whether any differences exist between factual and fictional behavior that affect the tasks of prediction and explanation. I then ask in what way causal accounts go beyond mere category judgments. Finally, I ask whether outcome biases affect both prediction and explanation.

EXPERIMENTATION, PREDICTION, AND EXPLANATION

> *"Most researchers are aware that randomized experiments are considered the 'gold standard' for causal inference."* – Rohrer (2018, p. 27)

Academic psychology is modern in the sense that one of its core goals is to uncover general laws governing mind and behavior, and in that its principal instrument is the experiment (Wundt, 1874; Woodworth and Schlosberg, 1954). With the search for laws, and theoretical explanations of these laws, psychological science aims to attain a fuller understanding of the nature of mental life (Popper, 1962; Meehl, 1978). In the modernist understanding, natural laws are there to be discovered, described, and deployed. At the vanishing, or "omega," point, science would reveal "how the mind works" and "where behavior comes from." At that limit, there would be a comprehensive ability to predict and explain mental states and behavior. This modernist understanding is rather mechanistic and deterministic, and it may seem outdated.[2] Postmodernism, much like quantum physics (Schrödinger, 1930), questions the validity of this paradigm, but it is not evident that postmodernism can ground itself in quantum theory. Nor that it would want to. To a committed postmodernist, Schrödinger's equations are just another story. However, most modernist scientists recognize the limits of determinism and the irreducibility of deep uncertainty. Still, the modernist premise that human experience is in large part comprehensible and lawful implies that experimental research is a powerful, if probabilistic, way to attain some "explanation through prediction."

Theories are prediction machines, and experiments are their testing grounds (see Hempel and Oppenheim, 1948;

[1] Tellingly, Copernicans are not known as "Postptolemaians."

[2] See Busemeyer et al. (2011) for post-mechanistic foundations of cognitive psychology.

Cummins, 2000; Grosz et al., 2020;, for critical discussions).[3] In the experimental context, theory and testing look to the future. Experimenters might well assert that an intervention "explains" a certain amount of the variance in the data, and that prediction proper is limited to the domain of machine learning, computational modeling (Yarkoni and Westfall, 2017) and mind-free behaviorism (Skinner, 1981). The view I take here is that experiments test theoretically grounded explanations (Deutsch, 2016; Holtz and Monnerjahn, 2017). Predictions, if corroborated by the data, can aid explanation. The data convey information about the hypothesis that predicted them, a hypothesis that now lies in the past (the current preference for pre-registration demands that it does). Inferences from the data back to the hypothesis, that is, to the statement that predicted the data, are "reverse," and they are fraught with uncertainty (Krueger and Heck, 2017, 2018).

Many people, when first introduced to the science of psychology, declare that they too have a keen interest in understanding what happened. "I am always curious," they say, "why people do what they do." As this desire to know springs from a natural epistemic instinct, it can be unsettling to learn that experimental psychologists have little to say about this question. Their inability to explain a particular behavior better than the folk themselves comes as a painful realization. In this article, I intend to take a small step toward a clarification of the nature of the disconnect between prediction and explanation by providing examples showing that findings obtained in the context of prediction can improve explanations of past events.

Postmodernism, at least in its early deconstructionist stage, rejected the idea of objective and knowable laws (Pluckrose and Lindsay, 2020). Gergen (1973) in particular denied that social behavior is lawful. He argued that culture, time, and individual reactions against presumptive laws erode any predictable regularities researchers might wish to consider lawful. Indeed, Gergen claimed that in the social domain any regularities that might aspire to the status of law are self-nullifying. To overcome these difficulties, Gergen (2001) endeavored to reimagine social psychology along postmodern lines, but he did not provide clear guidance of what to do next (Krueger, 2002). Whether no attempts to explain human behavior can rise above the *narratological* or the *perfomative* only a postmodernist can know. I suspect, though, that most postmodernists would endorse the view that humans share a need to gain self-understanding. From this assumption it appears to follow that a credible story that makes sense of specific events or phenomena is a universal desideratum (Spoehr and Spoehr, 1994; Mukharji and Zeckhauser, 2020). In contrast, there is no obligation to shed light on what will happen next. At least in theory, humans have the option to assert – as Aischylos did – that the future is unknowable. With these caveats, I accept the sovereignty of experimental psychology and ask whether we can clarify the relationship between prediction and explanation. If we can, we might be able to mitigate the apparent irreconcilability between modern and postmodern psychology.

THE FUTURE AND THE PAST

Kierkegaard (1843) observed that "the philosophers say that life must be understood backwards. But they forget the other proposition, that it must be lived forwards." The ordinary person, as it were, walks through life facing backward. Events and experiences come into view only when they begin to recede into the past. This person understands that looking forward and looking backward present unique challenges, and they may wonder how these two perspectives might be related to each other. Is, for example, a retrospective causal analysis and explanation necessary for making predictions about the future? Can predictions, if they turn out to be accurate, inform explanations of events that had occurred before the predictions were tested?

If there were no relationship between prediction and explanation, how might explanation be grounded? Perhaps explanations are worthwhile in their own right even if they cannot improve predictions. An appeal an explanation's intrinsic value must steer clear of tautology. What is gained if all we can say that we like a good story because it is compelling? Perhaps a good story provides meaning, and the attainment of meaning can enhance subjective well-being (Baumeister, 2018). On this view, finding meaning is a pleasure not unlike the gratification of a physiological drive (Chater and Loewenstein, 2016). This pragmatic justification of the explanatory project requires more empirical study. Even if positive correlations between perceived meaning and subjective well-being are found (Ho et al., 2010), there are grounds for skepticism. Some explanations may be accepted uncritically simply because they are plausible. Many conspiracy theories provide no testable predictions, and the explanations they offer are factually false (Douglas et al., 2017). Yet, such theories can offer an illusory sense of meaning and understanding (Forgas and Baumeister, 2019).

Why would anyone seek valid predictions if these predictions do not contribute to an understanding of the past? To be sure, the search for successful predictions is valuable because instrumental rationality demands that choices depend on future prospects and not past outcomes (Dawes, 1988; Krueger, 2000). If predictions and their outcomes also improve an understanding of the past, this is an added benefit. There should be some convergence between prediction and explanation if there is any truth to the idea that Nature is lawful.

In a complex world, it is difficult to separate signals from noise. Prediction and explanation are hard, but they are hard for different reasons. Consider the reasons that might be offered for the claim that one is harder than the other. Advocates of prediction may note the intrinsic uncertainty of the future (Prigogine, 1997). Making predictions is risky, and risk and uncertainty are fundamentally aversive (Frenkel-Brunswik, 1949; Gigerenzer, 2014; Krueger and Grüning, in press; but see Hertwig and Engel, 2020, for significant exceptions). Data may refute the hypothesis. Errors are great teachers, but they inflict

[3]Cummins (2000) argues that psychology has had greater success predicting regularities in human performance in experiments than explaining this performance in lawful terms. Many regularities, as for example, those summarized by psycho-physical laws, are descriptive instead of explanatory.

pain. Advocates of prediction may assert that past-oriented explanations are easier because the past has provided data that are now on the table, ready to be investigated. Investigators can look for probative evidence until they have reached a threshold at which they are willing to consider an explanation sufficient (Pennington and Hastie, 1992). A rewarding state of certainty is attained, and there is no immediate fear of failure.[4] Finding an explanation is not a bet. The past is always there to be explained or reinterpreted, but the future continually wastes away and turns into the past. Once an event occurred, we cannot go back and predict it again.[5]

Conversely, advocates of explanation may note that their work is harder because of the complexities of causality. There are usually many potential causes to explain an event, affording many possible stories and interpretations. The explainer has to distinguish between necessary and sufficient causes, and decide how many of each should be part of the explanatory account (Kelley, 1972a,b). Explanation is hard because the potential causes can only be fitted to the data; they cannot be tested, which would require prediction and new data. Use of thought experiments with counterfactuals is one tool to evaluate the fidelity of a causal account, but then again, counterfactual causes and effects are only that – imaginary (Norton, 1996). The subjective satisfactoriness of a causal account is a poor guide to its validity.[6] Someone might present a better account, but it is not clear what it would take for an account to fail in the kind of decisive way in which a prediction can fail.

In short, both advocates of prediction and advocates of explanation may claim that either approach is the harder one, depending on their epistemic or rhetorical aims. If there is a difference in difficulty, one wonders if it is an essential one or one that merely reflects differences in knowledge or differences in what advocates wish to emphasize. Russell (1913) once argued that there is no essential difference between predication and causation, and that these terms should be "extruded" from the philosophical vocabulary. Assuming perfect determinism in the tradition of Spinoza or Laplace, Russell argued that the direction of the flow of time has no bearing on how events are related to one another or how contingent they are on others. An omniscient being could wind the universe forward or backward, and the deterministic laws would reveal themselves in the same fashion. Requiring only the capacity of perception, this omniscient being would have no need to "think" in the way ordinary humans do when struggling to make a prediction or find an explanation. The need for prediction and explanation is a function of human ignorance. Of course, Russell's hypothetical bird's-eye view is a metaphysical *amuse bouche* (Prigogine, 1997). It should, however, remind the predictors and the explainers that they are looking at the same Nature, albeit from different angles.

Experimental psychology is dedicated to the study of causes and the "explanation of variance" with a forward-looking logic. Hypotheses are bets about data not yet seen. A potential cause is activated in the laboratory and its effects are observed. Yet, past-oriented explainers want a causal account of things that already happened. To the experimentalist, the question is this: Once a cause C is found to be sufficient to produce effect E, how can the presence of C be inferred once E has come into evidence? In other words, when predictors and explainers converse, they can take the problem of reverse inference as their common ground (Krueger, 2017b).

INFERENCES IN THE LAB AND IN THE WILD

By formalizing the relationships among unconditional and conditional probabilities, Bayes's Theorem provides a framework for thinking about reverse inference. Before reviewing the theorem and some of its implications, consider a thought experiment to illustrate the divergent interests of predictors and explainers. An experimenter has proposed the following hypothesis: "Men who want power admire men who have power." This sounds simple enough and perhaps tautological, but it is a prediction that might be worth testing. The experimenter measures the need for power in each of a number of sampled male participants. The experimenter then randomly assigns the participants to an experimental and a control condition. The men in the experimental condition receive a treatment – perhaps by way of priming or persuasion – designed to temporarily increase their need for power. The experimenter then measures the admiration these participants express for certain high- vs. low-power men.

Suppose that, as hypothesized, the experiment shows that the experimental manipulation did not affect admiration for low-powered men but did produce a strong effect on the admiration for high-powered men. Suppose for simplicity that the data are normally distributed within each condition and the difference between the two means is one standard deviation. The experimenter can now ask how probable it is that a randomly drawn participant had received the experimental treatment. In this symmetrical case, forward and reverse inferences yield the same result. The probability of a participant with a score above the grand average to belong to the treatment group is the same as the probability of a participant from the treatment group to have a score above the grand average. With $d = 1.0$, this probability is 0.69. An effect size of $d = 0.4$, which is empirically more realistic, yields a rather modest categorization benefit of $p = 0.58$, with $p = 0.5$ being the floor of ignorance.

This weakness of reverse inference for a typical effect size highlights a critical feature of experimentation: the narrow focus on group averages (Danziger, 1994). An individual's score is modeled as the sum of the group average and an "error" term, which comprises both the imperfections of measurement and whatever it is that makes the individual unique (Lord, 1959). The laws pursued by experiments sampling participants do not exhaust all that Nature has to offer; they are limited to group trends. If, in the above example, the difference between the two group means were to remain the same while the variance

[4] Although perhaps there should be because of the risk of "explanation bias" (Mukharji and Zeckhauser, 2020).

[5] An additional difficulty of prediction is that it requires the prediction of the causes of events in order to predict the events themselves.

[6] Bowers and Davis (2012) present a pointed critique of "just-so stories" in cognitive science, and Bayesian theories in particular.

of the individual's scores increased within each condition, the standardized effect size would shrink and reverse inferences would become even more uncertain.

Inferences after experimentation tend to underestimate the true effect. In the hypothetical experiment, the need of power was measured first, but it was not used to predict the admiration for powerful men. Individual differences on this measure were treated as error variance. If there is is a correlation between need for power and admiration of the powerful within conditions, it seems likely that this correlation contains a causal effect of need on admiration. This situation is analogous to one where a correlation between "need" and "love" is observed in the wild. When an experiment shows that manipulated need predicts heightened love, this finding affords the inference that the original, non-experimental, correlation contains a causal path from need to admiration.[7]

Experimental results understate the difficulty of making causal inferences in the wild. In the wild, many potential causes remain in play. Whereas the skillful experimenter eliminates uninteresting causes *a priori*, the skillful explainer must be an expert explorer. The task is to detect and eliminate improbable causes after the fact. In a world of many causes and many effects, the scenarios faced by the predictor and by the explainer look different. The predictor is interested in one cause, C, realizing that it may have several effects, E_1 to E_k. Being interested in this one cause and in only one particular effect, say E_1, the experimenter has attempted to manipulate only this one cause and to neutralize all others by randomly assigning the participants to the experimental and the control conditions; the other potential effects are dismissed by not being measured. The task is to see if C predicts E_1. In contrast, the explainer has selected a particular event or effect E, and wants to know which of several potential causes, C_1 to C_k, is the most effective one. Where can the predictor and the explainer meet?

BAYESIAN REVERSE INFERENCE

Bayes's Theorem shows that the probability of a cause C given an effect E is equal to the product of the prior probability of C and the "diagnostic ratio," which is the probability of the effect E given cause C over the unconditional probability of E, or

$$p(C|E) = p(C) \cdot \frac{p(E|C)}{p(E)} \quad (1)$$

A one-to-one association between experimental results, which yield p(E|C), and what the explainer wants, which is p(C|E), is limited to the special case in which the cause is as probable as the effect *a priori*, that is, if p(C) = p(E). Such an equality is rare. The probabilities of C and E are often unknown, but there are regularities that provide common ground for predictors and explainers.

Consider first the implications of the prior probability of the effect. This probability is equal to the sum of the products of the unconditional probabilities of the various causes in play and their corresponding conditional probabilities of yielding the effect, or

$$p(E) = p(C_1) \cdot p(E|C_1) + \ldots p(C_k) \cdot p(E|C_k) \quad (2)$$

To understand the implications of this equation, consider the case where all causes are ineffectual, that is, they neither promote nor inhibit the effect. All diagnostic ratios, that is, all $p(E|C_i)/p(E) = 1$. There are three important implications. First, once one cause is found to have a diagnostic ratio >1, the diagnostic ratios of all other causes fall below 1. As one cause is identified as promoting the effect, all others must now be assumed to be inhibitors. Second, p(E) increases if at least one $p(E|C_i) > 1$, while all other $p(E|C_i)$ remain the same. That is, the price of having found some relevant causal information is that the effect is less rare than formerly thought. Third, as more causes of the promoting kind emerge, they reduce the number of inhibitory causes, and p(E) increases further while the diagnostic ratio of each individual promoting cause becomes smaller. Likewise, the inverse conditional probability, $p(C_i|E)$, for each promoting cause also becomes smaller, although their sum increases.

Consider a numerical example. Causes C_1 to C_4 each have a prior probability of 0.25. The first cause is perfectly promoting, $p(E|C_1) = 1$, while the other three are perfectly inhibiting, $p(E|C_{2\text{ to }4}) = 0$. Now, p(E) = 0.25, the diagnostic ratio is 4 for C_1 and 0 for the three others, and the inverse conditional, p(C|E) is 1 for C_1 and 0 for the three others. Next, we assume that C_1, C_2, and C_3 are found to be maximally promoting, that is, $p(E|C_1) = p(E|C_2) = p(E|C_3) = 1$. The result is that p(E) rises to 0.75, while the diagnostic ratios of the promoting causes fall to 1.33, and their probabilities given the effect fall to 0.333. For any particular number of presumed causes, the more frequent the effect is (the higher p(E) is), the less effective individual causes are. Frequent events are thus difficult to explain with a parsimonious account, that is, an account that requires few causes. The more promoting causes there are, the more common the effect is likely to be and the weaker is the role for each individual cause. By contrast, rare events are potentially well explained by few – perhaps even just one – highly effective causes.[8] Conversely, common effects are easy to predict. One need only bet on common events of the past to repeat themselves (Hull et al., 1947; Ouellette and Wood, 1998). Often, the predictors of common effects are not even referred to as causes, but simply as "conditions," or general states of nature prevailing before and after the appearance of the effect. In contrast, rare effects are difficult to predict (Taleb, 2007; but see Lindaas and Pettersen, 2016). Black swan events are infamous for not having been predicted but then having been explained *ad libitum*.

Now consider the consequences of variation in the prior probability of a cause. The more probable a cause is, the less likely it is to be highly effective. This is so because the cause's effectiveness is captured by the diagnostic ratio of $p(E|C_i)/p(E)$

[7]In trivariate correlational models, correlations between presumed mediators and criterion variables are easily mistaken as support for a causal claim (Fiedler et al., 2011), a claim that can, however, be corroborated by experiments manipulating the mediator (Spencer et al., 2005).

[8]This Bayesian analysis is limited to sufficient causes, which are assumed to be mutually exclusive and exhaustive. Interactions among them are not being considered.

and because $p(C_i)$ is part of the denominator (see Eq. 2). A rare cause that explains an effect that would otherwise not occur is most attractive. A compelling explanatory account reveals how an unusual or extraordinary event came about thanks to the force of a single cause that would otherwise rarely be seen. Many of humankind's legends and myths comprise rare causes stirring up extraordinary outcomes. Achilles gets mad only twice, when Agamemnon steals his concubine Briseis and when Hector kills his friend Patroclus. In both cases, Achilles responds promptly, with sullen retreat and mortal rage respectively after the theft and after the murder. Good stories are memorable because they provide a crisp causal account (Schank and Abelson, 1977). Vivid one-to-one cause-and-effect associations do not require laborious probability calculations; they are open to "direct perception" (Heider and Simmel, 1944), particularly if they obey Hume's contiguity requirement in time and space. The explanatory causal account forces itself upon the observer. But had these events and their consequences been predictable? Our myths and legends are thick on causal stories, but thin on predictions. There is the occasional dark prophecy, which usually goes unheeded, or, as in Aischylos's Prometheus, there is "blind hope." Conjunctions of rare causes and rare events are the pleasures of the explanatory mind (a temptation that I will indulge in the second section of this article), but a challenge to those who seek to make successful predictions. Yet, many experimental psychologists wish to demonstrate causal relationships that seem unlikely at the outset for fear of having demonstrated what turns out to be trivial, tautological, or familiar in folk psychology (Felin et al., 2019). Researchers must "anticipate the unexpected," and do so without making it seem paradoxical (Fiedler, 2017). The more experimentalists pursue the high-hanging fruits of risky hypotheses, the more often they will fail, thereby stoking the discipline's replication crisis. But when they succeed, they may be able to make non-trivial contributions that are attractive to students of history (Krueger and Heck, 2018).

HISTORY INSPIRES RESEARCH WHICH HELPS TO EXPLAIN HISTORY

I now turn to the question of whether reverse inferences from theory and experimental findings can help explain past behavioral episodes. Can psychological research shed light on events that are otherwise the reserve of historical analysis or folk psychological interpretation? Note that the question of whether such linkages can be found is different from the mission and scope of applied social psychology, which seeks to generalize interventions that have been found to work in the lab. Like basic theory-driven work, applied social psychology is concerned with the optimization of future outcomes, that is, with making predictions (Forgas et al., 2020).

As to the influence of the wild on the lab, the history of social psychology is instructive. Early social psychological research advanced in part by responding to social problems such as Anti-Semitism (Adorno et al., 1950; Martin, 2001), other ethnic stereotypes QQ (Katz and Braly, 1933; Krueger, 1996a,b), excessive conformity (Asch, 1956; Constant et al., 2016), or yielding to propaganda (Hovland et al., 1953; Osterhouse and Brock, 1970; Lewandowsky et al., 2013). Two iconic research programs owed their existence to specific historical events and the expectation that experimental results would shed light on why the actors behaved as they did. Arendt's (1963) account of the Eichmann trial stimulated Milgram's (1963) obedience studies. Arendt suggested that obedience to authority is a sufficient cause of abhorrent behavior, and is perhaps the preponderant cause. Milgram sought to show that the essential dynamics, that is, the lawful regularities, of authority and obedience can be reproduced in the laboratory. During that same decade, the murder of Kitty Genovese prompted Latané and Darley (1968) to experimentally reproduce the phenomenon of bystander apathy. These real-world events were striking and were considered improbable at the time; yet they could be demonstrated by experimental research. For the purposes of this article, it is worth noting that reverse inferences from the data could extend beyond the hypotheses tested in the laboratory to retrograde explanations of the historical events that had inspired the research. What is more, when there are new instances of violence in a state-sponsored context or failures to intervene on behalf of others in need, the findings of experimental psychology contribute to the construction of causal accounts.

Today, a problem-focused approach to research continues (Krueger and Funder, 2004) with comparatively little attention paid to applications, interventions, or historical analysis. Many researchers focus on theory development and theory testing, where applications or reverse inferences are left to others. Much progress has been made in theory development and theory testing, although it is not always clear what is meant by "theory." Rigorous hypothetico-deductive processes are not the rule in experimental design (Prager et al., 2018), and many researchers settle for vague verbal descriptions of phenomena (Gigerenzer, 1998, 2010).

The three historical cases, or case studies, described below are meant to illustrate how research evidence obtained in the laboratory as well as theoretical analysis can inform historical case studies. Intended not as proofs of concept but rather as suggestions of concept, these vignettes show that theory-grounded research can make useful contributions to historical analysis. The first vignette seeks to improve the understanding of a man's fateful personal decision not only with reference to the historical context but also in light of cognitive research on decision-making. The second vignette seeks to show how an extreme case of identity transformation may be understood in light of theory and research on escalating commitments. The third vignette uses a contemporary theory of strategic behavior in addition to experimental findings to reconstruct an interpersonal dynamic and its – as it turns out – benign resolution.

PHILIPP VON HUTTEN: MAKING FATEFUL DECISIONS UNDER THE SHADOW OF DEATH

During the first half of the 16th Century, German bankers financed several expeditions into what was then known as "Little Venice," or Venezuela. The goal of these expeditions was to

subjugate the native population and to exploit its natural and cultural resources. Little is remembered today about this chapter of American history. Hence, the travelog and the letters of Philip von Hutten are of great interest (Schmitt and von Hutten, 1996). Hutten's activities and experiences can be seen through a politico-historical lens, with an emphasis on the European project to conquer and colonize the Americas. There is nothing in the present analysis to detract from this approach. The question of interest here is whether the available information sheds light on Hutten's psychology, and whether contemporary research on judgment and decision making can help illuminate his fateful decision to carry on (Krueger, 2013).

The critical event is Hutten's decision to mount a third expedition into the Venezuelan hinterland when two prior multi-year expeditions had already failed to bear fruit. Hutten was deep in debt, needing to replenish supplies without having made enough plunder to cover the costs, and it had become clear that the prospects of finding riches were remote. His family in Germany seems to have understood his dire circumstances. They offered to pay his debts if he returned home. Hutten refused, noting the unacceptability of the ridicule he was sure to suffer at the hands of his German neighbors. This is a psychological element of note, as it expresses the ethics of honor that was standard for men of Hutten's caste (Cohen et al., 1996). In social psychological terms, Hutten was concerned with the image he would project and the judgments he would receive in the social world to which he would return (Krueger et al., 2020).

Is Hutten's fear of being judged a sufficient explanation? Although such fears can be debilitating, few prefer death to ridicule. The record does not suggest that Hutten was suicidal. He was aware of the mortal risks he was facing, but he did not foresee that he would be captured and beheaded by a Spaniard who was a rival for the position of governor. The critical question is how he evaluated the chances of a third expedition to finally yield the badly needed rewards. A first pass is to submit that Hutten was overconfident in predicting success. Excessive confidence precipitates more failures than successes in life and business (Moore, 2020).

Beyond overconfidence lies the possibility of escalating commitments culminating in a sunk cost fallacy (Arkes and Blumer, 1985; Arkes and Ayton, 1999; Feldman and Wong, 2018). Commitments escalate as each successive investment of time, money, honor, or other material or psychological resources makes additional investments more likely regardless of real or imagined profits. Research on escalating commitments itself was inspired by famous failures in business and war. For example, the protraction of the Vietnam war beyond the point of its apparent failure presented a challenge to researchers to model, predict, and ultimately understand patently irrational behavior. Why, or under what conditions, are some individuals willing to do what they don't want to do – and even when they have an exit option? Why, in economic terms, would people pay for something they are loath to do?

Hutten's own documents show that he knew his previous expeditions were failures, that he knew his planned expedition had a low expected value of success (although it cannot be proven that he thought this value was negative), and that he knew that his family was ready to cover his debts. In reconstruction, the confluence of the ethic of honor along with the pull of sunk costs as documented by experimental research, provide a sufficiently plausible explanation for his fateful decision to persist (Krueger, 2013)

GONZALO GUERRERO: TRANSFORMING IDENTITY ONE STEP AT A TIME

A few years before Hutten, another European adventurer's life took an unexpected turn. Gonzalo, also known by the honorific appellation "Guerrero," was not a nobleman or would-be conquistador, but a mariner sailing with a Spanish expedition to Panama. The attempt to build a colony there failed. The Spaniards decided to return to Cuba, but their ship sank off the coast of the Yucatán. The survivors were captured by local Maya who proceeded to eat all but two of them. Gonzalo and a padre named of Jerónimo de Águilar were spared to be consumed later. The two escaped and found refuge with a rival tribe, were enslaved again, but evaded mortal threat. This is the beginning of a legend as told by the chroniclers (de Landa, 1566) and contemporary Gonzalo scholars (Calder, 2017). Their common source was a single eyewitness, namely Águilar. Perhaps Gonzalo never existed. His story could be a myth. For the present project, this does not matter. His story raises questions about what humans are capable of doing *in extremis*, and whether experimental psychology can help explain how.

Gonzalo not only survived but flourished. In time, he won his captors' trust and respect, married a chief's daughter, begot children, and became an influential war captain. Along the way, he became more Maya in his thinking, feeling, and acting. The outward signs of his transformation are critical. In the Yucatán today, where he is revered as an ancestral figure, he is depicted with tattoos, piercings, and various native ornamentations. Yet, statues and paintings also show him with a beard to note his origins (Mueller, 2001).

If Hutten escalated his commitments toward a fateful business and life decision, Gonzalo went down a road of stepwise identity transformation (Krueger, 2017a). By what he did, he changed who he was. It is not clear whether Gonzalo's transformation comprised sunk costs. A demonstration of sunk costs requires evidence of costs outpacing rewards, while stopping remains possible. Gonzalo himself, however, claimed that his life had improved. According to Águilar, he declined to rejoin the Spaniards when given the chance. Pointing to his family and the social status he had achieved, Gonzalo asked Águilar if it made sense for him to forsake it all. The life he had built was a rewarding one, and his decision to cultivate it further might well have been rational.

Although the Gonzalo legend shares with Hutten's story the mechanisms of escalating commitments (Burger, 1999; Krueger and Massey, 2009), it shows how – under the right circumstances – such commitments can yield extraordinary results. As a thought experiment, the Gonzalo legend explores the limits of identity transformation, a process millions of

immigrants and other travelers confront every day (Hong et al., 2016). By asking what is possible, Gonzalo poses a special challenge to experimental research staked on the prediction of averages in the search for general laws.[9]

One line of empirical research that can contribute to an understanding of the Gonzalo experience is work on world-class performance. The late K. Anders Ericsson showed in a multi-decade research program that sustained practice at the edge of one's ability eventually yields the rewards of world-class performance (Ericsson, 2016). The effects of sustained and challenging practice on the attainment of expertise were first demonstrated in narrowly defined performance domains such as playing chess or playing a musical instrument (Ericsson et al., 1993). The constraints of a narrowly defined space of performance are crucially important. Loosely or ill-defined domains such as "leadership ability" or "wisdom of life" do not allow the model to work (Krueger, 2020).

A person's project to assimilate into a new culture entails a broader set of skills than playing chess, but it is still more clearly defined than "being wise." Hence, Ericsson's theory of expertise applies *mutatis mutandis*. The mastery of a formerly alien culture requires the acquisition of a new language, a code of conduct, an understanding of norms and expectations, along with the acquisition of other forms of "tacit knowledge" (Polanyi, 1966). The dedicated immigrant is drawn into multiple correlated expertise projects, which can be fueled further by irreversible commitments (Schelling, 1956; Koziel et al., 2010). In short, Gonzalo's life is beginning to make sense in light of theory-grounded psychological science – without losing its charm.[10]

ROBINSON CRUSOE: FINDING SOCIAL PEACE BUT NO EQUILIBRIUM

In his masterly novel, Defoe (1719/1998) explored a man's response to extraordinary circumstances. Shipwrecked, Robinson Crusoe solves the problems of survival without human company, only to discover that the presence of other humans brings new challenges, which are in some ways more daunting. Driven more by fear and prudence than by aggression or greed, Crusoe kills a party of cannibals, and saves the life of one of their victims. He names him Friday, as it was on a Friday that he found him. Crusoe is an even poorer fit to the prototype of the conquistador than are Hutten or Gonzalo. Yet, Crusoe has guns, while Friday has none; and Crusoe saved Friday's life, although he can't be sure whether Friday will remember the deed with gratitude.

The psychological questions are "How did Crusoe and Friday manage to get along?" and "Why did they not kill each other?" Some social scientists have recognized the intellectual and theoretical appeal of this puzzle (Tsebelis, 1989; van Lange et al., 2014), but have presented no analysis. I attempted one

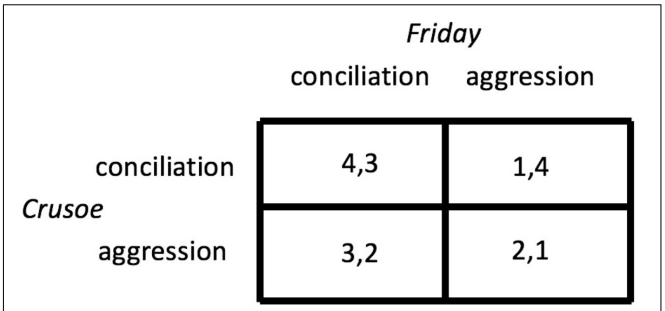


FIGURE 1 | The power game in matrix form. The row player's (Crusoe's) payoffs are shown to the left of the comma. Higher numbers are better.

(Krueger, 2014a,b), which I summarize here. This analysis is theory-driven and it can be enriched with experimental results.

The theory is the heterodox game "theory of moves" (Brams, 1993, 2011). This theory asks the analyst to generate a plausible rank ordering of the four outcomes pertaining to each interactant. These payoffs arise from the crossing of the two strategies each player has: to be aggressive or to be conciliatory. Relying on my reading of Defoe's account, I suggest the following rank order for Crusoe. Crusoe's primary interest is to have a conciliatory (i.e., submissive) Friday, and if Friday submits, Crusoe is happy to be conciliatory as well. If, however, Friday is aggressive (i.e., rebellious), Crusoe would rather fight than flee. Friday, for his part, is primarily interested in Crusoe – who has guns – to be conciliatory. But if Crusoe is indeed gentle, Friday has an incentive to rebel. Otherwise, Friday will submit to an aggressive Crusoe (again, Crusoe has guns).

With the two sets of preferences in place (shown in **Figure 1**), a game theoretic analysis looks for a player's best response assuming the other player's preferences are known. Both Crusoe and Friday face a tricky dilemma because neither has a dominating strategy and the game has no unique Nash equilibrium.[11] The best collective outcome is obtained if both players are conciliatory, and this is how Defoe tells it. This leaves Crusoe with his best (4) and Friday with his second-best outcome (3). This may seem unfair, but given the player's misaligned preferences all joint outcomes are unfair.

The theory of moves raises interesting questions, two of which I will address. The first question is whether Crusoe is more powerful than Friday, and if so, how this advantage is encoded in the preference rankings. Crusoe's doing better than Friday is an indication of his greater power, but we cannot consider it sufficient without courting an outcome bias. Crusoe's advantage in outcome is no clear signal of his greater power causing this favorable result. Crusoe's and Friday's primary interests are not diagnostic either. Both prefer the other to be conciliatory. Their secondary interests are different, however. Playing tit-for-tat, Crusoe claims the power to reward conciliation and to punish

[9] Hirschman (2013) noted that, compared with the physical sciences, the social sciences remain sensitive to the exploration and extreme and unique cases, an orientation he calls "possibilism."

[10] It is well to remember that Jerómino de Águilar outlived Gonzalo without taking the path of full acculturation. Águilar's success suggests that Gonzalo had some choice in the matter, as an alternative survival strategy was available.

[11] A strategy (aggression or conciliation) is dominating if it is the best response irrespective of the other player's move. A unique Nash equilibrium is a combination of strategies that leaves neither player with an incentive to act differently.

rebellion. Friday does the opposite. He would exploit conciliation and yield to aggression. If Friday also played tit-for-tat, the game would be a "stag hunt," where mutual cooperation is easily achieved (Krueger et al., 2020). In other words, it is Friday's own interest in power that makes Crusoe more powerful.

The second question is how a state of mutual conciliation can be maintained if it is not a Nash equilibrium. Here we see that it depends on Friday. Friday has an incentive to rebel, moving the game to "Crusoe 1: Friday 4." Yet, Crusoe would counter by putting the rebellion down, yielding "Crusoe 2: Friday 1." An intelligent player can foresee how this game would deteriorate into a cycle of war and truce, leaving average outcomes to both players that are inferior to the ones obtained with mutual conciliation. Viewed from this perspective, Friday is the wiser player and perhaps the more powerful one. He controls the keeping of the peace. The price he pays is the cost of his deference to Crusoe.

Findings from experimental social psychology and behavioral economics offer further insights. Social psychological research supports the notion that humans are sensitive to hierarchy and reluctant to challenge power when such challenges are risky and costly (Koski et al., 2015). Once established, challenges to power hierarchies have a better chance of succeeding if the challengers form alliances (DeScioli and Kurzban, 2009). Single challengers are at the greatest risk of failing. A behavioral economics perspective notes the similarity between the final outcome of "Crusoe 4: Friday 3" and the typical result of an ultimatum game (Güth, 1995). Crusoe claims his first preference and asks Friday to settle for something short of his, Friday's, maximum. Most participants in ultimatum games accept such a positive if mildly unequal distribution. Taken together, the present *post hoc* analysis of Defoe's fiction shows that to the psychologist, Friday is the more interesting character.[12] Some postmodernists will agree, and one author re-told the Crusoe story from Friday's perspective. In his account, "*Vendredi*" introduces Crusoe to the superior ways of living untarnished by the European Enlightenment (Tournier, 1969). Supposing that one fiction cannot disprove another, I settle for showing that psychological science can help explain behavior in rational terms.[13]

DISCUSSION

My attempt to make theoretical and experimental (social) psychology useful for the interpretation of observed (or imagined) behavior is only a sketch. There is no smooth path from theory and experiment to a sufficient explanation of historical (or fictional) events. In the first section of this article, I reviewed the logic and the challenges of reverse inference in rather abstract fashion. In the second section, I illustrated the potential of theory and research to help explain real and imagined historic events. To conclude, I discuss three issues that complicate this enterprise, although I do think that the challenges are surmountable. I begin with the question of whether behavior, in order to be explained, must be "real." Then I ask whether, as a reverse inference, a causal account of behavior is different from other types of category judgment. Finally, I explore to what extent outcome bias can affect both predictions and explanations.

FICTION, WHAT FICTION?

Moving from Hutten to Gonzalo to Crusoe, we descended from reasonably well-documented history to legend to pure fiction. I have suggested that the methods of causal analysis apply regardless of where the events of interest are located on this spectrum. How might this claim be justified? One response is that no such justification is possible. On this view, real and imagined behavioral episodes are different kinds. They differ from one another much like lived experience differs from dreams. If so, it is dangerous to ignore this difference in kind and to treat it as if it were a matter of degree. Philosopher Nozick (1981) argued that people, as a rule, prefer fact to fiction, no matter how pleasing this fiction might be. Ironically, Nozick made this point by way of a thought experiment. A thought experiment is fiction, and fiction can simulate and explain reality by bringing into focus critical similarities and differences. All counterfactuals are *per definitionem* fiction, but they are useful in causal analysis (Byrne, 2016). This assumption allows the factual and the counterfactual to switch roles. If the counterfactual can help explain the factual, so can the factual help explain the counterfactual.

There are parallels between lived events ("factual behavior") and imagined events ("fictive behavior"). Both are entangled in the causal web of the world. Fiction that violates the constraints of this web strikes us as bizarre or entertaining as "science" fiction. Fiction becomes "relatable" inasmuch as it enables readers to construct a causal story that makes sense of what happens. Doing so, they perform the same cognitive operations they would perform for real events. A causal analysis of a piece of fiction, if it is informed by findings obtained with future-oriented, that is, predictive, psychological science, is comparable to a causal analysis of a fact. Both analyses are simulations,[14] whose goal it is to provide a satisfactory explanatory account, or a "story," with an acceptable goodness of fit. Such an account cannot be refuted in the same way that a prediction can be refuted by data. A good causal story can only be replaced by a better one, if such a story comes along.

CAUSES AS CATEGORIES

I have argued that future-oriented and prediction-based science can aid past-oriented and explanation-based scholarship and lay cognition. Bayes's Theorem served as an orienting framework. Inferences from the future to the past are reverse inferences.

[12] Arguably, the title of the essay "How Robinson Crusoe managed his Man Friday" (Krueger, 2014a) was a misnomer. Friday's strategic choices, it turns out, were more pivotal for the preservation of collectively beneficial outcomes. Friday, in other words, managed Crusoe.

[13] Even if the *explanandum* itself turns out to be irrational – as in the Hutten case.

[14] Whether performed on fact or fiction, causal accounts are simulations in the sense that causation is, in either case, a psychological moment added to represent the relationship of what is then called the cause and the effect (De Pierris and Friedman, 2018).

They tend to work, but without more information or strong assumptions, it is difficult to know just how well they work; one only knows that making such inferences works better than doing nothing (Krueger and Heck, 2017).

A skeptical view is that there is little that is new because the benefits and limitations of reverse inference are well known (Dawes, 1988; Krueger, 2017b). An implication of this sort of skepticism is that causal inference reduces to category inference and that therefore the cognitive errors corrupting the latter also corrupt the former. This claim has merit inasmuch as one is willing to assume that anything that can be modeled with Bayes's Theorem is of the same kind. In psychology (Kahneman and Tversky, 1973) and in medical science (Eddy, 1982), processes of category judgment have attracted a great deal of attention (Fiedler and Grüning, in press). The signature finding is the so-called base-rate fallacy, which occurs when people place an instance into a small category if the probability of that instance is high *given* membership in that category (see Gigerenzer and Hoffrage, 1995, or Koehler, 1996, for critical evaluations). The base-rate fallacy is a one-to-one reverse inference that ignores the prior probability of the category. In the medical context, this fallacy entails overdiagnoses of rare diseases. Even many trained diagnosticians confuse a test's sensitivity with its positive predictive value (Franklin and Krueger, 2003). People, to paraphrase Robyn Dawes, assume a symmetry Nature does not provide (see Krueger, 1996b, for other violations of Bayesian inference).

From a Bayesian perspective, one might assert that causes are categories just like other categories. This view can be psychologized by saying that once a sufficient cause has been found for the event of interest, this event is placed into the category comprising all those events that result from this particular cause. The usual psychological biases would occur, and the base rate fallacy is of particular concern. The fallacy would be present if, after observing a high p(E|C), the perceiver concluded that p(C|E) is also high even if p(C) is low. In other words, the behavior would be attributed to a potent cause too rare to emerge as the most probable.

If causes were just a kind of category, the work of inference would be simple. Arguably, however, there is a difference. A complete causal inference has two elements. The identification of a probable cause where the resulting p(C|E) passes some threshold, say 0.5, is only the first element. The second element is the identification of a causal process. In Bayes's Theorem, this process is represented by the extent of belief updating, which is captured by various ratios or difference scores [e.g., p(E|C)/p(E) or p(C|E) – p(C)]. The most convincing causal claims comprise both a causal categorization where p(C|E) is high and a causal process where p(C|E) – p(C) is large. The two elements are conceptually independent, though statistically related (Krueger and Heck, 2017). It is the latter element, the causal process, that gives energy to the human tendency to convey lessons of causality in story form.[15]

[15]With regard to usage, I note that many Bayesians refer to inferences about categories as "predictions." Indeed, the conditional probability of p(C|E) is also known as the "positive predictive value." In contrast, I see causal inferences in the temporal context of judgments about the past, reserving the term "prediction" for inferences about events that have not yet occurred.

OUTCOME BIAS IN INFERENCES ABOUT THE FUTURE AND THE PAST

Along with overconfidence, confirmation bias, innumeracy, and downright foolishness, outcome bias is one of the signature threats to rational reasoning. Baron and Hershey (1988) made the canonical case, and the bias has been documented many times since. The bias occurs when the outcome of a decision contaminates evaluations of decision quality. When a decision maker takes a calculated risk or makes a decision under uncertainty, the evaluator should limit the assessment of decision quality to the information the decision maker had (or failed to obtain) and to the processing of this information. Information the decision maker did not have or could not have at decision time ought to be excluded from the assessment. The outcome of a decision, by definition, follows a decision, and must therefore be ignored (Krueger and Acevedo, 2007). A lottery winner may be congratulated for having had good luck but should not be praised for having the gift of prophecy; a loser should not be blamed for losing. If anything, both should be blamed for playing in the first place as lotteries have steeply negative expected values.

The concept of outcome bias straddles the tasks of prediction and explanation. The decision maker is engaged in a prediction task, and the observer evaluates how well this task is performed. To do this, the evaluator simulates the process of making the prediction. Outcome bias occurs if the evaluator bestows too much praise on the decision maker after a positive outcome, or too much blame after a negative outcome (as noted in the lottery example above). At this level, outcome bias is unrelated to the question of *why* the decision makers made a particular prediction.

How might outcome bias affect explanations and causal inferences? The booming field of moral psychology has dedicated itself to the question of how people assign blame to transgressors. It was once believed that an outcome bias exists such that the weight of a negative outcome directly predicts the degree of assigned blame regardless of other considerations such as the transgressor's mental state at the time (Mazzocco et al., 2004; Cushman, 2008). This view is no longer viable (see Malle et al., 2014, for an extended review and discussion). Malle et al.'s (2014) "theory of blame" stresses the role of mental state inferences and especially inferences about intentionality. Intentional actions tend to be supported or justified by reasons, which in turn predict judgments of blame. Crucially, this theory and others in the field of moral psychology take the presence (or absence) of personal causation as a starting point to get to judgments of blame. These theories are not concerned with the question of what caused the person to act beyond the intention immediately preceding the act. A more comprehensive causal explanation would also address the origins of intentions. If we were to say that Hutten planned a third expedition because he wanted to, we would be making a circular argument (Greve, 2001); likewise for the question of why Gonzalo became a Maya, or for the question of why Friday did not rebel against Crusoe.

It would be a severe case of outcome bias to infer intentions only from outcomes. Early attribution theorists (Heider, 1958; Jones and Davis, 1965) warned against this heuristic. In the present analysis of the three historico-literary cases, I took the

actors' intentions for granted, and asked about their underlying causes. It remains possible that outcome biases compromise such causal inferences. One might wonder, for example, if Hutten's tragic end at the hands of a murderer influenced my conclusion that his decision to undertake a third expedition was rationally flawed. It should not.

CONCLUSION

In the spirit of the arguments presented in this article I should ask "Why did I write this article?" There are several coincidental factors. The first and central of these factors is a longstanding concern with the question of how, as an experimental social psychologist, I can respond to those who ask how my discipline enables us to explain people's behavior. To me, the question became "whether" my discipline produces answers. Recognizing the difficulty of the problem, coupled with the general apathy of the field with regard to this question, I experienced cognitive dissonance (Festinger, 1957). Finding myself unable to ignore the question and thinking I have learned enough to attempt an answer, I embarked on this writing project. In the spirit of looking for theoretically and empirically grounded explanations, I submit that my work has been motivated, at least in part, by a wish to justify my career-long investments in my profession (Alicke and Sedikides, 2011).

A second factor is a growing involvement with other disciplines, especially the humanities. The discussions of Hutten, Gonzalo, and the Crusoe-Friday team grew out of presentations given at international conferences dedicated to the legacy of Alexander von Humboldt, the champion of the universality and interconnectedness of all science. Each of the three stories has its own history and logic of having been selected at the time. Once selected, these stories were highly accessible to me. No claim of their representativeness is made. Relatedly, I discovered Steven Brams's "theory of moves" when studying "scripture," a form of literature. It was a perspective-shifting experience to see Brams (1980) apply his variant of game theory to the interpretation of myths and stories lying at the foundation of our civilization. That this could be done was, as it were, a revelation. Integrating the use of Brams's tools with social psychological theory and research to explain historical or literary events then came naturally. Again, in cognitive-psychological terms, these tools had become chronically accessible (Higgins, 1996).

A final question is whether the present effort can make a contribution to the further development of psychological science. Satisfying the folk by presenting compelling explanations of past behavior is one thing; improving theory and practice is another. The present article contains no new normative recommendations but an invitation to see that the possible uses of social psychology have been underestimated.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

The author did all the work on this article. Literature review, conceptualization, writing, preparation of the figure.

ACKNOWLEDGMENTS

Rosamel Benavides-Garb opened the gates to the humanities for me, and Simone Ebben, Tony Evans, David Grüning, Emily Siff, and David Urschler made stimulating and constructive comments on drafts of this article. I also gratefully acknowledge the reviewers' contributions to the final version of this article.

REFERENCES

Adams St. Pierre, E. (2002). "Science" rejects postmodernism. *Educ. Res.* 31, 25–27. doi: 10.3102/0013189X031008025

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., and Sanford, R. N. (1950). *The Authoritarian Personality*. Manhattan, NY: Harper and Row.

Alicke, M. D., and Sedikides, C. (eds) (2011). *Handbook of Self-Enhancement(and)Self-Protection*. New York, NY: Guilford Press. doi: 10.1093/oxfordhb/9780195399820.013.0017

Arendt, H. (1963). *Eichmann in Jerusalem: A Report on the Banality of Evil*. New York, NY: Viking.

Arkes, H. R., and Ayton, P. (1999). The sunk cost and concorde effects: are humans less rational than lower animals? *Psychol. Bull.* 125, 591–600. doi: 10.1037/0033-2909.125.5.591

Arkes, H. R., and Blumer, C. (1985). The psychology of sunk cost. *Organ. Behav. Hum. Decis. Process.* 35, 124–140. doi: 10.1016/0749-5978(85)90049-4

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.* 70, 1–70. doi: 10.1037/h0093718

Baron, J., and Hershey, J. C. (1988). Outcome bias in decision evaluation. *J. Pers. Soc. Psychol.* 54, 569–579. doi: 10.1037//0022-3514.54.4.569

Baumeister, R. F. (2018). "Happiness and meaningfulness as two different and not entirely compatible versions of the good life," in *The Social Psychology of Living Well*, eds J. P. Forgas and R. F. Baumeister (Abingdon: Routledge), 21–33. doi: 10.4324/9781351189712-2

Bereiter, C. (1994). Implications of postmodernism for science, or, science as progressive discourse. *Educ. Psychol.* 29, 3–12. doi: 10.1207/s15326985ep2901_1

Bowers, J. S., and Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychol. Bull.* 138, 389–414. doi: 10.1037/a0026450

Brams, S. J. (1980). *Biblical Games. A Strategic Analysis of Stories in the Old Testament*. Cambridge, MA: MIT Press.

Brams, S. J. (1993). Theory of moves. *Am. Sci.* 81, 562–570. doi: 10.1017/CBO9780511558979

Brams, S. J. (2011). *Game Theory and the Humanities*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262015226.001.0001

Burger, J. M. (1999). The foot-in-the-door compliance procedure: a multiple-process analysis and review. *Pers. Soc. Psychol. Rev.* 3, 303–325. doi: 10.1207/s15327957pspr0304_2

Busemeyer, J. R., Pothos, E. M., Franco, R., and Trueblood, J. S. (2011). A quantum theoretical explanation for probability judgment errors. *Psychol. Rev.* 118, 193–218. doi: 10.1037/a0022542

Byrne, R. (2016). Counterfactual thought. *Annu. Rev. Psychol.* 67, 135–157. doi: 10.1146/annurev-psych-122414-033249

Calder, R. (2017). *A Hero for the Americas: The Legend of Gonzalo Guerrero*. Regina: University of Regina Press.

Chater, N., and Loewenstein, G. (2016). The under-appreciated drive for sense-making. *J. Econ. Behav. Organ.* 126, 137–154. doi: 10.1016/j.jebo.2015.10.016

Cohen, D., Nisbett, R. E., Bowdle, B. F., and Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: an "experimental ethnography.". *J. Pers. Soc. Psychol.* 70, 945–960. doi: 10.1037/0022-3514.70.5.945

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., and Friston, K. (2016). Regimes of expectations: an active inference model of social conformity and human decision making. *Front. Psychol.* 10:679. doi: 10.3389/fpsyg.2019.00679

Cummins, R. (2000). "'How does it work?' versus 'What are the laws?': Two conceptions of psychological explanation," in *Explanation and Cognition*, eds F. Keil and R. Wilson (Cambridge, MA: MIT Press), 117–144.

Cushman, F. (2008). Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition* 108, 353–380. doi: 10.1016/j.cognition.2008.03.006

Danziger, K. (1994). *Constructing the Subject*. Cambridge: Cambridge University Press.

Dawes, R. M. (1988). *Rational Choice in an Uncertain World*. San Diego, CA: Harcourt Brace Jovanovich.

de Landa, D. (1566). *Relación de Las Cosas de Yucatán*, 2nd Edn. Mérida, YU: Royal Academy of History.

De Pierris, G., and Friedman, M. (2018). "Kant and Hume on causality," in *The Stanford Encyclopedia of Philosophy*, (Winter 2018 Edition) ed. E. N. Zalta (Stanford, CA: CSLI). .

Defoe, D. (1719/1998). *Robinson Crusoe*. Illinois, III: Dover.

DeScioli, P., and Kurzban, R. (2009). The alliance hypothesis in human friendship. *PLoS One* 4:e5802. doi: 10.1371/journal.pone.0005802

Deutsch, D. (2016). The logic of experimental tests, particularly of Everettian quantum theory. *Stud. Hist. Philos. Phys.* 55, 24–33. doi: 10.1016/j.shpsb.2016.06.001

Dörner, D., and Güss, C. D. (2011). A psychological analysis of Adolf Hitler's decision making as commander in chief: summa confidentia et nimius metus. *Rev. Gen. Psychol.* 15, 37–49. doi: 10.1037/a0022375

Douglas, K. M., Sutton, R. M., and Cichocka, A. (2017). The psychology of conspiracy theories. *Curr. Dir. Psychol. Sci.* 26, 538–542. doi: 10.1177/0963721417718261

Eddy, D. M. (1982). "Probabilistic reasoning in clinical medicine: problems and opportunities," in *Judgment under Uncertainty: Heuristics and Biases*, eds D. Kahneman, P. Slovic, and A. Tversky (Cambridge: Cambridge University Press), 249–267. doi: 10.1017/CBO9780511809477.019

Ericsson, K. A. (2016). *Peak: How to Master Almost Anything*. London: Penguin.

Ericsson, K. A., Krampe, R. T., and Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychol. Rev.* 100, 363–406. doi: 10.1037/0033-295X.100.3.363

Feldman, G., and Wong, K. F. E. (2018). When action-inaction framing leads to higher escalation of commitment: a new inaction-effect perspective on the sunk-cost fallacy. *Psychol. Sci.* 29, 537–548. doi: 10.1177/0956797617739368

Felin, T., Felin, M., Krueger, J. I., and Koenderink, J. (2019). On surprise-hacking. *Perception* 48, 109–114. doi: 10.1177/0301006618822217

Festinger, L. (1957). *A theory of Cognitive Dissonance*. Indiana, TX: Row, Peterson and Company.

Feyerabend, P. (1975). *Against Method: Outline of an Anarchistic Theory of Knowledge*. London: New Left Books.

Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspect. Psychol. Sci.* 12, 46–61. doi: 10.1177/1745691616654458

Fiedler, K., and Grüning, D. J. (in press). A social psychological toolbox for clinical psychology. *Z. Psychol.*

Fiedler, K., Schott, M., and Meiser, T. (2011). What mediation analysis can (not) do. *J. Exp. Soc. Psychol.* 47, 1231–1236. doi: 10.1016/j.jesp.2011.05.007

Forgas, J. P., and Baumeister, R. F. (2019). "The social psychology of gullibility: fake news, conspiracy theories, and irrational beliefs," in *Proceedings of the Sydney Symposium of Social Psychology*, Vol. 21, (London: Psychology Press). doi: 10.4324/9780429203787

Forgas, J. P., Crano, W. D., and Fiedler, K. (2020). Applications of social psychology: how social psychology can contribute to the solution of real-world problems. *Proceedings of the Sydney Symposium of Social Psychology*, Vol. 22. London: Psychology Press. doi: 10.4324/9780367816407

Franklin, R. D., and Krueger, J. (2003). "Bayesian inference and belief networks," in *Prediction in Forensic and Neuropsychology: Sound Statistical Methods*, ed. R. D. Franklin (Mahwah, NJ: Erlbaum), 65–87. doi: 10.4324/9781410607089

Frenkel-Brunswik, E. (1949). Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.* 18, 108–143. doi: 10.1111/j.1467-6494.1949.tb01236.x

Gergen, K. J. (1973). Social psychology as history. *J. Pers. Soc. Psychol.* 26, 309–320. doi: 10.1037/h0034436

Gergen, K. J. (2001). Psychological science in a postmodern context. *Am. Psychol.* 56, 803–813. doi: 10.1037/0003-066X.56.10.803

Gigerenzer, G. (1998). Surrogates for theories. *Theory Psychol.* 8, 195–204. doi: 10.1177/0959354398082006

Gigerenzer, G. (2010). Personal reflections on theory and psychology. *Theory Psychol.* 20, 733–743. doi: 10.1177/0959354310378184

Gigerenzer, G. (2014). *Risk Savvy*. London: Penguin.

Gigerenzer, G., and Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: frequency formats. *Psychol. Rev.* 102, 684–704. doi: 10.1037/0033-295X.102.4.684

Greve, W. (2001). Traps and gaps in action explanation: theoretical problems of a psychology of human action. *Psychol. Rev.* 108, 435–451. doi: 10.1037/0033-295X.108.2.435

Grosz, M. P., Rohrer, J. M., and Thoemmes, F. (2020). The taboo against explicit causal inference in nonexperimental psychology. *Perspect. Psychol. Sci.* 15, 1243–1255. doi: 10.1177/1745691620921521

Güth, W. (1995). On ultimatum bargaining experiments—A personal review. *J. Econ. Behav. Organ.* 27, 329–344. doi: 10.1016/0167-2681(94)00071-L

Heider, F. (1958). *The Psychology of Interpersonal Relations*. New York, NY: Wiley. doi: 10.1037/10628-000

Heider, F., and Simmel, M. (1944). An experimental study of apparent behavior. *Am. J. Psychol.* 57, 243–259. doi: 10.2307/1416950

Hempel, C., and Oppenheim, P. (1948). Studies in the logic of explanation. *Philos. Sci.* 15, 135–175. doi: 10.1086/286983

Hertwig, R., and Engel, C. (2020). *Deliberate Ignorance: Choosing not to Know*. Cambridge, MA: MIT Press.

Higgins, E. T. (1996). "Knowledge activation: accessibility, applicability, and salience," in *Social Psychology: Handbook of Basic Principles*, eds E. T. Higgins and A. W. Kruglanski (New York, NY: Guilford Press), 173–198.

Hirschman, A. O. (2013). "Political economy and possibilism," in *The Essential Hirschman*, ed. J. Adelman (Princeton, NJ: Princeton University Press), 1–34. doi: 10.1515/9781400848409-002

Ho, M. Y., Cheung, F. M., and Cheung, S. F. (2010). The role of meaning in life and optimism in promoting well-being. *Pers. Individ. Dif.* 48, 658–663. doi: 10.1016/j.paid.2010.01.008

Holtz, P., and Monnerjahn, P. (2017). Falsificationism is not just 'potential' falsifiability, but requires 'actual' falsification: social psychology, critical rationalism, and progress in science. *J. Theory Soc. Behav.* 47, 348–362. doi: 10.1111/jtsb.12134

Hong, Y.-Y., Zhan, S., Morris, M. W., and Benet-Martínez, V. (2016). Multicultural identity processes. *Curr. Opin. Psychol.* 8, 49–53. doi: 10.1016/j.copsyc.2015.09.020

Hovland, C. I., Janis, I. L., and Kelley, H. H. (1953). *Communication and Persuasion*. London: Yale University Press.

Hull, C. L., Felsinger, J. M., Gladstone, A. I., and Yamaguchi, H. G. (1947). A proposed quantification of habit strength. *Psychol. Rev.* 54, 237–254. doi: 10.1037/h0062631

Hutchinson, I. (2011). *Monopolizing Knowledge*. Belmont, MA: Fias Publishing.

Jones, E. E., and Davis, K. E. (1965). From acts to dispositions: the attribution process in person perception. *Adv. Exp. Soc. Psychol.* 2, 219–266. doi: 10.1016/S0065-2601(08)60107-0

Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychol. Rev.* 80, 237–251. doi: 10.1037/h0034747

Katz, D., and Braly, K. (1933). Racial stereotypes on one hundred college students. *J. Abnorm. Soc. Psychol.* 28, 280–290. doi: 10.1037/h0074049

Kelley, H. H. (1972a). "Attribution and social interaction," in *Attribution: Perceiving the Causes of Behavior*, eds E. E. Jones, D. E. Kanouse, H. H. Kelley,

R. S. Nisbett, S. Valins, and B. Weiner (New York, NY: General Learning Press), 1–26.

Kelley, H. H. (1972b). "Causal schemata and the attribution process," in *Attribution: Perceiving the Causes of Behavior*, eds E. E. Jones, D. E. Kanouse, H. H. Kelley, R. S. Nisbett, S. Valins, and B. Weiner (New York, NY: General Learning Press), 151–174.

Kierkegaard, S. (1843). *Journals VII 1A 363*.

Koehler, J. J. (1996). The base rate fallacy reconsidered: descriptive, normative, and methodological challenges. *Behav. Brain Sci.* 19, 1–17. doi: 10.1017/S0140525X00041157

Koski, J. E., Xie, H., and Olson, I. R. (2015). Understanding social hierarchies: the neural and psychological foundations of status perception. *Soc. Neurosci.* 10, 527–550. doi: 10.1080/17470919.2015.1013223

Koziel, S., Kretschmer, W., and Pawlowski, B. (2010). Tattoo and piercings as signals of biological quality. *Evol. Hum. Behav.* 31, 187–192. doi: 10.1016/j.evolhumbehav.2009.09.009

Krueger, J. (1996a). Personal beliefs and cultural stereotypes about racial characteristics. *J. Pers. Soc. Psychol.* 71, 536–548. doi: 10.1037/0022-3514.71.3.536

Krueger, J. (1996b). Probabilistic national stereotypes. *Eur. J. Soc. Psychol.* 26, 961–980. doi: 10.1002/(SICI)1099-0992(199611)26:6<961::AID-EJSP799>3.0.CO;2-F

Krueger, J. (2000). Distributive judgments under uncertainty: Paccioli's game revisited. *J. Exp. Psychol.* 129, 546–558. doi: 10.1037/0096-3445.129.4.546

Krueger, J. I. (2002). Postmodern parlor games. *American Psychologist* 57, 461–462. doi: 10.1037/0003-066X.57.6-7.461

Krueger, J. I. (2013). *Psychology of a Conquistador. Jahrbuch für Europäische Überseegeschichte*, Vol. 13. Wiesbaden: Harrassowitz Verlag, 37–41.

Krueger, J. I. (2014b). *A General Power Game*. Available online at: https://www.psychologytoday.com/us/blog/one-among-many/201402/general-power-game (accessed July 1, 2020).

Krueger, J. I. (2014a). *How Robinson Crusoe Managed his Man Friday*. Available online at: https://www.psychologytoday.com/us/blog/one-among-many/201402/how-robinson-crusoe-managed-his-man-friday (accessed July 1, 2020).

Krueger, J. I. (2017a). *Going Native*. Available online at: https://www.psychologytoday.com/us/blog/one-among-many/201708/going-native (accessed July 1, 2020).

Krueger, J. I. (2017b). "Reverse inference," in *Psychological Science Under Scrutiny: Recent Challenges and Proposed Solutions*, eds S. O. Lilienfeld and I. D. Waldman (Hoboken, NJ: Wiley), 110–124. doi: 10.1002/9781119095910.ch7

Krueger, J. I. (2020). The phantom of the omniscient leader. *Am. J. Psychol.* 133, 414–418.

Krueger, J. I., and Acevedo, M. (2007). Perceptions of self and other in the prisoner's dilemma: outcome bias and evidential reasoning. *Am. J. Psychol.* 120, 593–618. doi: 10.2307/20445427

Krueger, J. I., and Funder, D. C. (2004). Towards a balanced social psychology: causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behav. Brain Sci.* 27, 313–327. doi: 10.1017/S0140525X04000081

Krueger, J. I., and Grüning, D. J. (in press). "Psychological perversities and populism," in *The Social Psychology of Populism. The 22nd Sydney Symposium on Social Psychology*, eds J. P. Forgas and K. Fiedler (Milton Park: Taylor & Francis).

Krueger, J. I., and Heck, P. R. (2017). The heuristic value of p in inductive statistical inference. *Front. Psychol.* 8:908. doi: 10.3389/fpsyg.2017.00908

Krueger, J. I., and Heck, P. R. (2018). Testing significance testing. *Collabra* 4:11. doi: 10.1525/collabra.108

Krueger, J. I., Heck, P. R., Evans, A. M., and DiDonato, T. E. (2020). Social game theory: preferences, perceptions, and choices. *Eur. Rev. Soc. Psychol.* 31, 322–353. doi: 10.1080/10463283.2020.1778249

Krueger, J. I., and Massey, A. L. (2009). A rational reconstruction of misbehavior. *Soc. Cogn.* 27, 785–810. doi: 10.1521/soco.2009.27.5.786

Latané, B., and Darley, J. M. (1968). Group inhibition of bystander intervention in emergencies. *J. Pers. Soci. Psychol.* 10, 215–221. doi: 10.1037/h0026570

Lewandowsky, S., Stritzke, W. G. K., Freund, A. M., Oberauer, K., and Krueger, J. I. (2013). Misinformation, disinformation, and violent conflict: from iraq and the "War on Terror" to future threats to peace. *Am. Psychol.* 68, 487–501. doi: 10.1037/a0034515

Lindaas, O. A., and Pettersen, K. A. (2016). Risk analysis and Black Swans: two strategies for de-blackening. *J. Risk Res.* 19, 1231–1245. doi: 10.1080/13669877.2016.1153499

Lord, F. M. (1959). An approach to mental test theory. *Psychometrika* 24, 283–302. doi: 10.1007/BF02289812

Malle, B. F., Guglielmo, S., and Monroe, A. E. (2014). A theory of blame. *Psychol. Inquiry* 25, 147–186. doi: 10.1080/1047840X.2014.877340

Mandel, D. R. (2014). The psychology of bayesian reasoning. *Front. Psychol.* 5:1144. doi: 10.3389/fpsyg.2014.01144

Martin, J. L. (2001). The authoritarian personality, 50 years later. what questions are there for political psychology? *Polit. Psychol.* 22, 1–26. doi: 10.1111/0162-895X.00223

Mazzocco, P. J., Alicke, M. D., and Davis, T. L. (2004). On the robustness of outcome bias: no constraint by prior culpability. *Basic Appl. Soc. Psychol.* 26, 131–146. doi: 10.1207/s15324834basp2602&3_3

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *J. Consult. Clin. Psychol.* 46, 806–834. doi: 10.1037/0022-006X.46.4.806

Milgram, S. (1963). Behavioral study of obedience. *J. Abnorm. Soc. Psychol.* 67, 371–378. doi: 10.1037/h0040525

Moore, D. A. (2020). *Perfectly Confident*. New York, NY: HarperCollins.

Mueller, R. A. M. (2001). "From cult to comics: the representation of gonzalo guerrero as cultural hero in mexican popular culture," in *A twice told tale: Reinventing the Encounter in Iberian/Iberian American Literature and Film*, eds S. J. Navarro and T. R. Young (Newark, DE: University of Delaware Press), 137–148.

Mukharji, A., and Zeckhauser, R. (2020). Bound to happen: explanation bias in historical analysis. *J. Appl. Hist.* 1, 5–27. doi: 10.1163/25895893-00101002

Norton, J. D. (1996). Are thought experiments just what you thought? *Can. J. Philos.* 26, 333–366. doi: 10.1080/00455091.1996.10717457

Nozick, R. (1981). *Essays on Anarchy, State, and Utopia*. Lanham, MD: Rowman & Littlefield.

Osterhouse, R. A., and Brock, T. C. (1970). Distraction increases yielding to propaganda by inhibiting counterarguing. *J. Pers. Soc. Psychol.* 15, 344–358. doi: 10.1037/h0029598

Ouellette, J. A., and Wood, W. (1998). Habit and intention in everyday life: the multiple processes by which past behavior predicts future behavior. *Psychol. Bull.* 124, 54–74. doi: 10.1037/0033-2909.124.1.54

Pennington, N., and Hastie, R. (1992). Explaining the evidence: tests of the story model for juror decision making. *J. Pers. Soc. Psychol.* 62, 189–206. doi: 10.1037/0022-3514.62.2.189

Pluckrose, H., and Lindsay, J. (2020). *Cynical Theories*. Durham: Pitchstone.

Polanyi, M. (1966). *The Tacit Dimension*. Abingdon: Routledge.

Popper, K. R. (1962). *Conjectures and Refutations*. New York, NY: Basic Books.

Prager, J., Krueger, J. I., and Fiedler, K. (2018). Towards a deeper understanding of impression formation: new insights gained from a cognitive-ecological analysis. *J. Pers. Soc. Psychol.* 115, 379–397. doi: 10.1037/pspa0000123

Prigogine, I. (1997). *The End of Certainty*. New York, NY: Simon & Schuster.

Rohrer, J. M. (2018). Thinking clearly about correlations and causation: graphical causal models for correlational data. *Adv. Methods Pract. Psychol. Sci.* 1, 27–42. doi: 10.1177/2515245917745629

Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society* 13, 1–26.

Schank, R., and Abelson, R. (1977). *Scripts, Plans, Goals, and Understanding*. Mahwah, NJ: Erlbaum.

Schelling, T. C. (1956). An essay on bargaining. *Am. Econ. Rev.* 46, 281–306.

Schmidt, J. (2002). The legacy of the enlightenment. *Phil. Literature* 26, 432–442. doi: 10.1353/phl.2003.0016

Schmitt, E., and von Hutten, F. K. (1996). *Das Gold der Neuen Welt: Die Papiere des Welser-Konquistadors und Generalkapitäns von Venezuela Philipp von Hutten 1534 – 1541 (The gold of the New World: The papers of the Welser conquistator and captain general of Venezuela Philipp von Hutten)*. Hildburghausen: Verlag Frankenschwelle.

Schrödinger, E. (1930). Zum heisenbergschen unschärfeprinzip. Sitzungsberichte der preussischen akademie der wissenschaften. *Physikalisch-mathematische Klasse* 14, 296–303.

Simonton, D. K. (1998). Historiometric methods in social psychology. *Eur. Rev. Soc. Psychol.* 9, 267–293. doi: 10.1080/14792779843000108

Skinner, B. F. (1981). Selection by consequences. *Science* 213, 501–504. doi: 10.1126/science.7244649

Spencer, S. J., Zanna, M. P., and Fong, G. T. (2005). Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes. *J. Pers. Soc. Psychol.* 89, 845–851. doi: 10.1037/0022-3514.89.6.845

Spoehr, K. T., and Spoehr, L. W. (1994). Learning to think historically. *Educ. Psychol.* 29, 71–77. doi: 10.1207/s15326985ep2902_2

Taleb, N. N. (2007). *The Black Swan*. London: Penguin.

Tournier, M. (1969). *Friday*, Transl. by. N. Denny (New York, NY: Pantheon). Originally published in French, 1967.

Tsebelis, G. (1989). The abuse of probability in political analysis: the robinson crusoe fallacy. *Am. Polit. Sci. Rev.* 83, 77–91. doi: 10.2307/1956435

van Lange, P. A. M., Balliet, D., Parks, C. D., and Van Vugt, M. (2014). *Social Dilemmas: The Psychology of Human Cooperation*. Oxford: Oxford University Press.

von Humboldt, A. (1845–1858). *Kosmos: Entwurf Einer Physischen Weltbeschreibung*. German: Cotta. doi: 10.5962/bhl.title.4717

Woodworth, R. S., and Schlosberg, H. (1954). *Experimental Psychology*. London: Methuen.

Wundt, W. M. (1874). *Grundzüge der Physiologischen Psychology [Elements of Physiological 5 Psychology]*. German: Engelmann.

Yarkoni, T., and Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* 12, 1100–1122. doi: 10.1177/1745691617693393

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 07 December 2020
doi: 10.3389/fpsyg.2020.596425

The Principle of Inversion: Why the Quantitative-Empirical Paradigm Cannot Serve as a Unifying Basis for Psychology as an Academic Discipline

Roland Mayrhofer[1]* *and Fabian Hutmacher*[1,2]

[1] Department of Psychology, University of Regensburg, Regensburg, Germany, [2] Human-Computer-Media Institute, University of Würzburg, Würzburg, Germany

Keywords: self-perception of psychology, principle of inversion, methods in psychology, operationalism, definitions of psychology

OPEN ACCESS

Edited by:
Barbara Hanfstingl,
University of Klagenfurt, Austria

Reviewed by:
Maarten Derksen,
University of Groningen, Netherlands

***Correspondence:**
Roland Mayrhofer
roland.mayrhofer@ur.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *19 August 2020*
Accepted: *10 November 2020*
Published: *07 December 2020*

Citation:
Mayrhofer R and Hutmacher F (2020) The Principle of Inversion: Why the Quantitative-Empirical Paradigm Cannot Serve as a Unifying Basis for Psychology as an Academic Discipline. Front. Psychol. 11:596425. doi: 10.3389/fpsyg.2020.596425

INTRODUCTION

In the English-speaking world, as well as in the international academic discourse and many other languages, the term "science" or appropriate translations refer only to a certain area of knowledge, namely the natural and the social sciences, thus excluding what is usually referred to as humanities or *Geisteswissenschaften* (Szostak, 2004).

The history and philosophy of science shows that the sciences rely heavily on the so-called Scientific Method, a set of theoretical and methodological principles which consists, in essence, of observing, formulating hypotheses and testing these hypotheses in experiments, in order to discover general laws. In doing so, knowledge is generated by relying on empirical evidence, which in turn expresses directly observable phenomena in terms of quantitative data. Excluding the extremely complex discussions about advantages, drawbacks, and alternatives to the Scientific Method (e.g., Gower, 1997; Nola and Sankey, 2014; Andersen and Hepburn, 2015), it is important to note that quantitative-empirical methods and thinking according to the Scientific Method dominate the sciences (Haig, 2014; Nola and Sankey, 2014; Sankey, 2014) and also psychology (Garber, 2019; Haig, 2019; Toomela, 2020).

By contrast, the humanities rely much less on quantitative—let alone experimental—methods, although these are employed nonetheless when appropriate. Instead of producing and gathering empirical data, the humanities characteristically approach their subject matter from a descriptive, interpretive, and hermeneutical understanding whose historical and comparative angles cannot be conveniently summarized by a single term (Watanabe, 2010; Bem and De Jong, 2013; Leezenberg and De Vries, 2019).

PSYCHOLOGY AS AN ACADEMIC DISCIPLINE AND THE PRINCIPLE OF INVERSION

Psychology is usually portrayed as the study of the human mind and behavior, although this nomenclature does not make it entirely clear what it actually designates.

In ancient Greek, "*psyche*" encompasses a variety of meanings (Claus, 1983) such as stream of air, breath of life, substance of life (in an ontological sense), spirit, mind, soul, personality, consciousness, self, or even ghost (of the dead). Although these terms describe a semantic field with the underlying term "life," this does not tell us what psychologists actually investigate. The

term "psychology," in the sense of "study of the *soul*," originated in the early modern era and was employed mostly for topics which would be categorized to be part of philosophy today. Nevertheless, the founders of modern academic psychology with its predominantly experimental and empirical orientation in the nineteenth century—most notably Wilhelm Wundt—continued to use this term. Further historical research (such as Russo Krauss, 2019) might explain why "*psyche*" was retained as term for their subject matter, thus clarifying the conceptual ideas behind actual research.

This heterogeneity of "*psyche*" is mirrored in the different psychological subdisciplines, such as cognitive, social, or biological psychology and the many branches of applied psychology, such as educational, organizational, or clinical psychology. And while mind, personality, consciousness, and self are familiar terms in psychology, it is clear that other aspects of the Greek "*psyche*" such as the physical properties of breath are not part of the discipline.

In short, psychology investigates many aspects of human existence—but then how does psychology differ from, say, anthropology, sociology, or history? Wherein lies the unity of psychology as academic discipline?

A widely-used textbook (Gerrig, 2012, p. 2) gives the following answer: Many psychologists seek answers to the fundamental question: "What is human nature?" This question is pursued by looking at processes that occur within individuals, and thus psychology is defined as the scientific study of the behavior of individuals and their mental processes (see also e.g., Lilienfeld et al., 2015; Myers and Dewall, 2015 for similar conceptualizations). However, the main elements of this train of thought—human nature, the individual in its entirety, and a scientific approach—reveal that it is not a trivial matter to state precisely what psychology is actually about.

First, it seems debatable that Gerrig's attempt to subsume the subject matter of psychology under the umbrella term "human nature" is really more precise than hazy concepts such as "*psyche*" or "soul" which contemporary psychology has dismissed as too vague (Haaga, 2004; Henriques, 2004; Lilienfeld, 2004). Whereas academic psychologists might argue that they are not interested in such a hazy concept but rather in specific topics such as emotions, neurobiology, or education, all these concepts revolve around the human mind and behavior. Therefore, psychology does indeed have some kind of common theme or center—but this center is so vague that it cannot act as focal point or provide the same clear framework as the subject matters of other disciplines. By contrast, physics is also very diverse, possibly even more so than psychology, but its subject matter is clearly defined as matter and the related phenomena of energy, space and time.

Second, Gerrig's assumption that the answer to the question about "human nature" can be found "by looking at processes that occur within individuals" is not self-evident. Simply put, focusing on the individual is problematic because many—if not all—of the individuals' intrinsic processes are inextricably intertwined with larger social, societal, or historical contexts (e.g., Agassi, 1977; Margolis, 1995, 2008). In other words, the behavior of individuals and their mental processes are shaped by outside contextual and societal factors, which may vary over time. The failure to take variability into consideration might underestimate the complexity of mental processes and give the impression that "human nature" is more hard-wired and less context-dependent than it actually is.

All of this shows that the exact subject matter of psychology is hard to pinpoint or to distinguish from other disciplines which also deal with behavior and mental processes, such as anthropology, history, cultural and literary studies, or philosophy. Nevertheless, Gerrig's definition contains an element which is crucial for the self-conception of psychology as discipline, namely the emphasis on "scientific study." Similarly, the APA Dictionary of Psychology explicitly emphasizes "observation, experimentation, testing, and analysis" (VandenBos, 2015, p. 860) as characteristic research methods, echoing the Scientific Method.

As early as 1983, Jüttemann pointed out that the common factor underlying the various branches and areas of research in psychology is not characterized *thematically* but rather by a common method and methodology, namely the rather strict—and sometimes dogmatic—adherence to the Scientific Method. He termed this "the principle of inversion": in other disciplines method and methodology are aligned with the respective subject matter, but in psychology this principle is inverted (see also Royce, 1961; Michell, 1997; Summers, 2012).

Jüttemann's astute observation has two interesting consequences: First, there is a stark contrast between the rather strict methodological requirements and the very broad and often hazy thematic content of psychological research. Therefore, Jüttemann concludes ironically, "everything done by psychologists who employ the nomological methodology in their research must count as academic psychology" (Jüttemann, 1983, p. 34, translated). Second, this modus operandi differs from other academic fields which either have clearly defined subject matters or employ much less rigid methodologies.

The Scientific Method originated long before the institutionalization of psychology as an academic discipline in the late nineteenth century. Moreover, it is closely associated with the natural sciences, especially physics. Thus, psychology is dominated by a method which is neither unique to nor was developed within the own academic field. By contrast, methods which were developed within and specifically for the framework of psychology such as psychoanalysis or introspection are relegated to the fringes of academic psychology.

Jüttemann argues that the rigid methodology has far-reaching consequences for the very nature of psychology and criticizes the resulting "research operationalism," i.e., the fact that the subject matters of certain areas—such as stress—are only represented by phenomena and procedures which conform to operationalizations according to the Scientific Method. But by reducing complex phenomena to easily quantifiable laboratory procedures the concepts in question—e.g., stress—lose their original meaning. In essence, this means that by solely using the Scientific Method to investigate psychological phenomena (such as stress) we do not learn much about these phenomena as such. Rather, we transform them into something else which can be quantified and measured, meaning that highly complex phenomena are simplified in order to

make them quantifiable (see also Hibberd, 2019; Mayrhofer et al., in press). "Stress" is a concrete example because the concept of stress "is deeply intertwined with the constituents of modern identity" (Hutmacher, 2019, p. 181) and therefore extremely complex, although research on stress highly relies on quantifiable parameters.

Whereas this operationalizing of psychological concepts might be appreciated as a more precise specification, it also goes hand in hand with a narrowing of real-life phenomena. It is possible that stripping complex phenomena down to their—supposedly!—bare bones reveals their core mechanisms. But more often we lose important aspects during this process, thereby missing the opportunity to understand something in its entirety. In other words, the professed exactitude and the desire to uncover the fundamental mechanisms of mental and behavioral phenomena by employing a quantitative-empirical paradigm within the very framework of the Scientific Method inadvertently misses important aspects.

To put it differently, it is by no means self-evident that studying internal—mental or behavioral—processes according to a certain predefined method will yield the desired results or that other methods might not provide more or a different kind of insight. Furthermore, even if we gain some knowledge by applying the Scientific Method, it is not an evident conclusion that this will also tell us something about human nature.

DISCUSSION

Where does this leave academic psychology? Thematically, psychology is a very colorful picture of different subject matters, whose interconnectedness is often rather tenuous and does not display a strong cohesiveness while circling around "human nature" as a hazy center of gravity. However, this vibrant mixture is hidden behind a veneer of uniformity, which manifests itself in the strict adherence to the quantitative-empirical method. This uniformity certainly conveys strength because of its methodological rigor and scientific respectability. However, simultaneously it hampers psychology by preventing it from exploring other avenues which might yield additional insight into mental and behavioral processes or even human nature.

The idea of a one-stop method is problematic for two reasons: First, psychology as a field is wide and diverse. Second, the specific mental and behavioral phenomena—such as stress—are hard to define precisely (Zagaria et al., 2020). Therefore, applying something seemingly precise such as the Scientific Method is inherently at odds with trying to understand such hard-to-grasp, complex phenomena. In short, the quantitative-empirical method cannot serve as a unifying basis for psychology as an academic discipline because it misses important dimensions of "human nature."

We believe that postmodern approaches, which were specifically developed to describe the complexities and ambiguities of modern societies, may offer a way out of this dilemma, although here we can only give a brief sketch: Postmodern approaches recognize and emphasize that a certain phenomenon may be understood by using different methods. Seemingly different phenomena and/or approaches often point into the same direction, although from different perspectives (e.g., Bertens, 1995; Sim, 2011; Aylesworth, 2015). This does not mean that there is no "truth" in psychology or that we cannot approach this truth (Holtz, 2020). Rather, the strength of a postmodern mindset lies in the ability to describe and to comprehend very complex phenomena without watering them down.

Therefore, such approaches will probably expand both the range and the explanatory power of psychology. As mental and behavioral processes tend to be innately fuzzy, any investigation of these phenomena must take this fuzzy nature into account. This is of course no plea to abandon the quantitative-empirical methods as they have revealed many interesting aspects of the *psyche*. The human mind and behavior are diverse—so why should our methods for investigating them not be equally diverse? Although largely outside the "scientific mainstream," there are other schools of thought in psychology which operate on the basis of different concepts of science, such as psychoanalysis (Bazan, 2018), humanistic (Warmoth, 1998), constructivist (Lincoln and Hoffman, 2019), or phenomenological (Langdridge, 2007) psychology. Taking their approaches seriously is likely to turn academic psychology into a vibrant generator of relevant knowledge and to spark more light to the enigma we term *psyche*.

AUTHOR CONTRIBUTIONS

RM and FH developed the idea for the article. RM drafted the manuscript. FH provided feedback and suggestions. Both authors approved the manuscript for submission.

REFERENCES

Agassi, J. (1977). *Towards a Rational Philosophical Anthropology*. The Hague: Martinus Nijhoff. doi: 10.1007/978-94-010-1095-5

Andersen, H., and Hepburn, B. (2015). "Scientific method," in *The Stanford Encyclopedia of Philosophy,* ed E. N. Zalta. Available online at: https://plato.stanford.edu/entries/scientific-method/

Aylesworth, G. (2015). "Postmodernism," in *The Stanford Encyclopedia of Philosophy,* ed E. N. Zalta. Available online at: https://plato.stanford.edu/entries/postmodernism/

Bazan, A. (2018). Psychoanalysis and academia: psychoanalysis at the crossroads between exact and human sciences. *Int. Forum Psychoanal.* 27, 90–97. doi: 10.1080/0803706X.2017.1392040

Bem, S., and De Jong, H. L. (2013). *Theoretical Issues in Psychology. An Introduction*. Los Angeles, CA: SAGE. doi: 10.4135/9781526402097

Bertens, H. (1995). *The Idea of the Postmodern. A History*. London; New York, NY: Routledge.

Claus, D. B. (1983). *Toward the Soul: An Inquiry into the Meaning of psyché Before Plato*. New Haven, CT: Yale University Press.

Garber, I. (2019). Trends in contemporary psychology. *J. Russian East Eur. Psychol.* 56, 271–295. doi: 10.1080/10610405.2019.1659698

Gerrig, R. J. (2012). *Psychology and Life. 20th Edn*. Boston, MA: Pearson.

Gower, B. (1997). *Scientific Method: An Historical and Philosophical Introduction*. London: Routledge.

Haaga, D. A. (2004). Defining psychology: what can it do for us? *J. Clin. Psychol.* 60, 1227–1229. doi: 10.1002/jclp.20063

Haig, B. D. (2014). *Investigating the Psychological World: Scientific Method in the Behavioral Sciences*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9780262027366.001.0001

Haig, B. D. (2019). The importance of scientific method for psychological science. *Psychol. Crime Law* 25, 527–541. doi: 10.1080/1068316X.2018.1557181

Henriques, G. R. (2004). Psychology defined. *J. Clin. Psychol.* 60, 1207–1221. doi: 10.1002/jclp.20061

Hibberd, F. J. (2019). What is scientific definition? *J. Mind Behav.* 40, 29–52.

Holtz, P. (2020). Does postmodernism really entail a disregard for the truth? Similarities and differences in postmodern and critical rationalist conceptualizations of truth, progress, and empirical research methods. *Front. Psychol.* 11:545959. doi: 10.3389/fpsyg.2020.545959

Hutmacher, F. (2019). On the janus-facedness of stress and modern life. *J. Theor. Philos. Psychol.* 39, 181–192. doi: 10.1037/teo0000113

Jüttemann, G. (ed.). (1983). "Psychologie am Scheideweg: Teilung oder Vervollständigung?," in *Psychologie in der Veränderung. Perspektiven für eine gegenstandsangemessenere Forschungspraxis* (Weinheim: Beltz), 30–65.

Langdridge, D. (2007). *Phenomenological Psychology: Theory, Research and Method*. Harlow: Pearson.

Leezenberg, M., and De Vries, G. (2019). *History and Philosophy of the Humanities: an Introduction*. Amsterdam: Amsterdam University Press. doi: 10.2307/j.ctvnp0j8d

Lilienfeld, S., Lynn, S. J., Namy, L., and Woolf, N. (2015). *Psychology: From Inquiry to Understanding*. Boston, MA: Pearson.

Lilienfeld, S. O. (2004). Defining psychology: is it worth the trouble? *J. Clin. Psychol.* 60, 1249–1253. doi: 10.1002/jclp.20067

Lincoln, J., and Hoffman, L. (2019). Toward an integration of constructivism and existential psychotherapy. *J. Construct. Psychol.* 32, 108–125. doi: 10.1080/10720537.2018.1461719

Margolis, J. (1995). *Historied Thought, Constructed World. A Conceptual Primer for the Turn of the Millennium*. Berkeley, CA: UCP.

Margolis, J. (2008). *The Arts and the Definition of the Human. Toward a Philosophical Anthropology*. Stanford, CA: SUP. doi: 10.1515/9780804769860

Mayrhofer, R., Kuhbandner, C., and Lindner, C. (in press). The practice of experimental psychology. An inevitbaly postmodern endeavor. *Front. Psychol.* doi: 10.3389/fpsyg.2020.612805

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *Br. J. Psychol.* 88, 355–383. doi: 10.1111/j.2044-8295.1997.tb02641.x

Myers, D. G., and Dewall, C. N. (2015). *Psychology, 11th Edn*. New York, NY: Worth Publishers.

Nola, R., and Sankey, H. (2014). *Theories of Scientific Method. An Introduction*. Stocksfield: Acumen. doi: 10.4324/9781315711959

Royce, J. R. (1961). Heretical thoughts on the definition of psychology. *Psychol. Rep.* 8, 11–14. doi: 10.2466/pr0.1961.8.1.11

Russo Krauss, C. (2019). "The spread of the new definition of psychology," in *Wundt, Avenarius, and Scientific Psychology,* ed C. Russo Krauss (Cham: Palgrave Macmillan), 59–111. doi: 10.1007/978-3-030-12637-7_4

Sankey, H. (2014). "Scientific method," in *The Routledge Companion to Philosophy of Science,* eds M. Curd and S. Psillos (London: Routledge), 280–290.

Sim, S. (2011). *The Routledge Companion to Postmodernism, 3rd Edn*. London; New York, NY: Routledge. doi: 10.4324/9780203813201

Summers, F. (2012). Psychoanalysis, the tyranny of objectivism, and the rebellion of the subjective. *Int. J. Appl. Psychoanal. Stud.* 9, 35–47. doi: 10.1002/aps.310

Szostak, R. (2004). *Classifying Science. Phenomena, Data, Theory, Method, Practice*. Dordrecht: Springer.

Toomela, A. (2020). Psychology today: still in denial, still outdated. *Integr. Psychol. Behav. Sci.* 54, 563–571. doi: 10.1007/s12124-020-09534-3

VandenBos, G. R. (2015). *APA Dictionary of Psychology, 2nd Edn*. Washington, DC: American Psychological Association. doi: 10.1037/14646-000

Warmoth, A. (1998). Humanistic psychology and humanistic social science. *Hum. Soc.* 22, 313–319. doi: 10.1177/016059769802200306

Watanabe, T. (2010). Metascientific foundations for pluralism in psychology. *N. Ideas Psychol.* 28, 253–262. doi: 10.1016/j.newideapsych.2009.09.019

Zagaria, A., Ando, A., and Zennaro, A. (2020). Psychology: a giant with feet of clay. *Integr. Psychol. Behav. Sci.* 54, 521–562. doi: 10.1007/s12124-020-09524-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 07 January 2021
doi: 10.3389/fpsyg.2020.609823

The Epistemic Imperialism of Science. Reinvigorating Early Critiques of Scientism

Lucas B. Mazur[1,2]*

[1] Jagiellonian University, Kraków, Poland, [2] Sigmund Freud Privatuniversität Berlin, Berlin, Germany

Positivism has had a tremendous impact on the development of the social sciences over the past two centuries. It has deeply influenced method and theory, and has seeped deeply into our broader understandings of the nature of the social sciences. Postmodernism has attempted to loosen the grip of positivism on our thinking, and while it has not been without its successes, postmodernism has worked more to deconstruct positivism than to construct something new in its place. Psychologists today perennially wrestle to find and retain their intellectual balance within the methodological, theoretical, and epistemological struggles between positivism and postmodernism. In the process, pre-postmodern criticisms of positivism have been largely forgotten. Although they remain deeply buried at the core of psychology, these early alternatives to positivism are rarely given explicit hearing today. The current piece explores some of the early critiques of positivism, particularly of its scientism, as well as early suggestions to tip the scales (back) in favor of *sapientia* ("wisdom"). This third option, largely overlooked within mainstream psychology, is of tremendous value today as it is both deconstructive and constructive relative to the shortcomings of positivism. It avoids the overly reductionistic "trivial order" of positivism, as well as the deeply unsatisfying and disorienting "barbaric vagueness" of postmodernism, while simultaneously embracing important core elements of both currents of thought.

Keywords: science, scientism, positivism, postmodernism, sapientia, wisdom, epistemology

OPEN ACCESS

Edited by:
Ulrich Dettweiler,
University of Stavanger, Norway

Reviewed by:
Tom Dickins,
Middlesex University, United Kingdom
Roland Pfister,
Julius Maximilian University of Würzburg, Germany

***Correspondence:**
Lucas B. Mazur
lucas.mazur@uj.edu.pl

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *24 September 2020*
Accepted: *20 November 2020*
Published: *07 January 2021*

Citation:
Mazur LB (2021) The Epistemic Imperialism of Science. Reinvigorating Early Critiques of Scientism. Front. Psychol. 11:609823. doi: 10.3389/fpsyg.2020.609823

INTRODUCTION

> Psychological science is – for pragmatic considerations – mostly oriented toward positivism, which is perceived as a modern but not postmodern epistemology. Postmodern psychology has not yet arrived in mainstream psychology. [. . .] How can psychology move on from modern toward a postmodern science, and is this really necessary or useful? (Frontiers in Psychology, n.d.)

These lines from the call for papers are helpful for framing the arguments made in the text below as they point to several important characteristics of contemporary psychology. It has been long-clear that, while *as a methodology* positivism is powerful, *as a philosophy* for the social sciences it is deeply problematic (Sheen, 1934/2019). By taking the limits of method to be the limits of knowledge, we have artificially narrowed our understanding of knowledge. By way of the processes of conceptualization and operationalization that grant it its power, positivistic epistemology artificially limits the person (in its *Menschenbild*) to the empirical and measurable, and disregards or even denies all else. Positivism in psychology has been linked with our "almost

neurotic need to be seen as scientific" and "to reject the subjective world" (Baker, 1992, p.13). The attempt to more thoroughly describe or define positivism has been called a task taken up by the "presumptuous and masochistic" (Caldwell, 1994, p.1), a tongue and cheek expression of a sentiment found in several classic explorations of positivism, such as that by Mill (1865/2005) and Kołakowski (1968). Kołakowski (1968) argues that all such attempts are at least partially arbitrary, but that this arbitrariness is unavoidable if we are to meaningfully work with the term at all. Such an approach recognizes that within the history of positivism, "repetitions can take deceptively different forms" (Tolman, 1992, p.8). Thus, following the lead of such earlier scholarship, we will use a broad, inclusive understanding of positivism, a school of thought generally involving phenomenalism and the rejection of metaphysics, nominalism anchored in (usually quantifiable) empirical data, claims to value-free objectivity, and claims that science can be thought of as a (largely singular) enterprise that develops over time and that represents not *a* way of knowing the world but as *the* way of doing so (Kołakowski, 1968). The exact nature of this "temper of mind" or "style of thinking, which as a rule is not dealt with by its adherents" (Kołakowski, 1968, p.vi), can be "better known through the enemies of that mode of thinking than through its friends" (Mill, 1865/2005, p.1). Thus, we stand to gain a better appreciation of positivism by examining one of its opponents, postmodernism.

Postmodernism can be broadly understood as an attempt to reemphasis precisely those elements of our lives that tend to be marginalized or overlooked within positivistic epistemology (Hicks, 2011). Nevertheless, despite the best efforts of postmodern thinkers to tear down positivism, despite the label of positivism being disavowed by most and even seen as "perverse" by many psychologists today (Smythe, 1992), and despite increasingly evident epistemological and practical shortcomings of positivistic epistemologies (e.g., as seen in the "replication crisis" in psychology), positivism remains on center stage in mainstream psychology, teetering between being a philosophy and a methodology (Huniche and Sørensen, 2019). While the label of positivism is rarely explicitly used today, and even less so as a form of self-identification, exploring its influence on current psychology is in no way to "kick a dead horse," as the expression goes, precisely because the "horse is far from dead. Positivist thinking is too powerful, even today, to go away by itself" (Tolman, 1992, p.7). Steinmetz (2005) colorfully expressed the continued presence of positivism within various disciplines of the human sciences, including psychology, as follows: "Despite repeated attempts by social theorists and researchers to drive a stake through the heart of the vampire, the disciplines continue to experience a positivistic haunting" (p.3; see also Tolman, 1992; Laudan, 1996). Thus, regardless of who we see as the hero of the story (a pertinent question itself these days given the popular love of vampire stories), we seem to be at an impasse.

Rather than focusing on the points of conflict between positivism and postmodernism, in the current piece it is argued that we should take a step back so as to examine earlier, pre-postmodern objections to positivism and that in doing so we may identify a path or paths out of the apparent stalemate; paths that include core elements of both positivism and postmodernism. Thus, the arguments presented below are not new. They are, in fact, quite old. It will be argued that the way forward is based on old questions and old insights, and thus, this piece is what we might call a reminiscence. The answer to the problems of positivism lies not in a greater embrace of postmodernism, but in a more thorough examination of our older intellectual, if not academic, roots that are critical of positivistic *philosophy*. Rather than addressing the shortcomings of positivistic epistemology (i.e., method-turned-philosophy) by means of postmodernism, the field would be better served by deeper, more consequential reflections on *sapientia*, a form of metaphysical wisdom "that has seemed to grow so weak in the modern era, as *homo sapiens* has waned and *homo sciens* has waxed" (Aeschliman, 1983, p.20). To borrow the language of Whitehead (cited in Aeschliman, 1983, p.69), this classic third option avoids the overly reductionistic "trivial order" of positivism, whereby meaningfulness is sacrificed for mathematical precision, as well as the deeply unsatisfying and disorienting "barbaric vagueness" of postmodernism, whereby meaningfulness is lost in a world of subjectivity and relativity. At the same time, *sapientia* nevertheless embraces important core elements of both currents of thought.

We begin by examining the rise of materialistic philosophies in the 17th and 18th century, which led to the spread of positivism in the 19th century. In these periods, science was in the process of breaking away from other avenues of knowledge (e.g., the arts, theology), even though many scientists themselves remained deeply committed to them (Harrison, 2015). As materiality and quantification increasingly replaced spirit and quality as the means of validity, method came to constitute a philosophy in itself (Sheen, 1934/2019). Over time, the belief in the value of scientific methodology became *scientism*, the belief in science not only as *a* philosophical school of thought but as *the* philosophical school of thought. We then briefly explore how attempts to free the social sciences from the reductionism of positivistic philosophy took the form of postmodernism. While asserting the greater complexities of our lived experiences than can be seen based on positivism alone, postmodernism is primarily a deconstructive, reactive process, unable to truly free us from positivistic reductionism. Postmodernism has highlighted the shortcomings of scientism, and fought tooth and nail against it, but it is unable to move us forward on its own two feet. It is for this reason, as reflected in the call for papers, that psychology remains largely wedded to a modern epistemology that is simply deeply unsatisfying; we have developed an epistemological Stockholm syndrome, whereby we claim to be fleeing from a philosophy we simultaneously actively profess. Finally, we examine a handful of pre-postmodern positions based on *sapientia* that have been buried under the weight of our predominantly positivistic worldview. These positions can be reemphasized in a psychology that asserts the wholeness of the human, including the material and quantifiable, but also those parts beyond the conceptual reach of the scientific method (Mazur and Watzlawik, 2016; Mazur, 2017). This third path asserts the power of science *as method*, while also *critically* and *cautiously* supporting the polyvalence and complexities of life highlighted in postmodern thought. By reasserting the primacy

of metaphysics over methods, *sapientia* promotes the power of science without overextending it into philosophy, and it thereby encourages the kind of fundamental, agentic judgment and discrimination that can allow us to benefit from the insights of postmodernism without fear of being consumed by its "barbaric vagueness."

THE EPISTEMIC IMPERIALISM OF SCIENCE AND THE RISE OF SCIENTISM

"The epistemic imperialism of science" is a phrase used by Harrison (2015, p.190) to describe how science has come to be the dominant arbiter of what counts as genuine knowledge in the modern West. The empire of science was built over the last three centuries, but the true reach of its empirical aspirations only became clear relatively recently. Despite how recent science is in its modern form, it has been argued that earlier schools of thought constitute forerunners of the notion of science-as-philosophy. For example, pelagianism, the assertion that humans can reach a state of perfection by their own means without the help of God, has been pointed to as an early forerunner of positivism (Sheen, 1934/2019). Within psychology itself, the work of Ferdinand Ueberwasser (1752–1812) has been pointed to as a positivist project predating Wundt's famous laboratory by roughly a century (Schwarz and Pfister, 2016). Such earlier intellectual roots aside, what we now perceive to be a fundamental and eternal conflict between scientific and metaphysical epistemologies (including "faith-based" epistemologies) is a relatively recent development, and it would have been completely foreign to most people, including practitioners of empirical science, in the 17th, 18th, and even into the 19th centuries (Hunter, 2000; Harrison, 2015). An important element in the development of this modern epistemic rift is the emergence of the modern notion of objectivity, whereby validity is thought to be determined on the basis of quantification and tool-based mediation between the subject-as-researcher and the object of inquiry (Daston and Galison, 2007; Huniche and Sørensen, 2019), claims that also variously appeared in psychology (Danziger, 1990; Baker, 1992). By means of our objectifying tools, we have increasingly broken up the world into smaller and smaller parts (Gozli and Deng, 2018), a process that has been accompanied by an increased fragmentation of academic and intellectual disciplines (Cornejo, 2017). Huxley (1937, p.276) wrote that "intensive specialization tends to reduce each branch of science to a condition almost approaching meaninglessness" and that there "are many men of science who are actually proud of this state of things. Specialized meaninglessness has come to be regarded, in certain circles, as a kind of hall-mark of true science." While the social sciences, including psychology, are often thought of as latecomers to this process, social scientists in fact played an important role in the transformation of scientific methodology into a scientific philosophy (i.e., scientism), as seen in the development of influential social theory (e.g., Auguste Comte) and even the development of formal methods of quantification and statistical testing (e.g., Francis Galton, Karl Pearson, and Charles Spearman; Bury, 1920; Sheen, 1934/2019; Scriven, 1966). Over time, science came to be increasingly thought of as fundamentally separate and independent from other epistemologies (Daston and Galison, 2007). Mainstream psychology came to primarily conduct *method-driven research* (Danziger, 1990; Huniche and Sørensen, 2019). Steinmetz (2005) has argued that not only does psychology have a close relationship with positivism but also some subfields such as social psychology are openly hostile to non-positivistic approaches. The degree to which positivism has been absorbed into psychology, he argues, makes it difficult to identify subfields that are truly free from it: "Even to locate a frontier between positivism and non-positivism in the fields concerned with the psychic, then, we are best advised to move beyond psychology into psychoanalysis" (p.12). Interestingly, this "frontier" implies that even within Freud's psychoanalysis, there is a considerable degree of positivism (Elliot, 2005). What is more, most (but certainly not all) criticisms of psychoanalysis are based on various forms of positivism. For example, while Karl Popper may have been a vocal critic of *logical positivism*, his argument that psychoanalysis is essentially unfalsifiable is itself a positivistic position, although he did not himself like that term (Steinmetz, 2005). Popper largely rejected an inductionistic approach to science (1960); hence, his clash with the Vienna Circle and with what is often thought of as classical positivism. Within his hypothetico–deductive model of science, his deductivistic realism even makes room for imagination, fantasy, metaphysics, or guesswork as the starting point of scientific investigation. However, for Popper, these positions are only tenable to the extent that they be translated into empirically falsifiable scientific theories (Maxwell, 2017); hence, his line of thinking, as well as various other forms of hypothesis testing and probabilistic claims within psychology can be thought of as falling into the broad camp of positivism (and thus, he was attacked as a positivist by the Frankfurter School). Similarly, Popper believed intertheoretical translations to be of importance for the development of science, as it was only in this way that a new theory could be shown to be more falsifiable, but less falsified, than its predecessors (Laudan, 1996). Popper's positivism can be seen all the more when contrasted with the thinking of Kuhn, whose "relativism about standards is the exact counterpoint of Popper's methodological conventionalism" (Laudan, 1996, p.16). Nevertheless, although we are making use of a broad, inclusive, and rather simplistic understanding of positivism in this short piece, it needs to be clearly acknowledged that, like postmodernism, positivism is not a singular, homogenous philosophy, but rather speaks to a very diverse set of positions and practices (Kołakowski, 1968; Feichtigner et al., 2018). Mill's (1865/2005, p.1) 19th century observation on the use of this label is further illustrative of this point: "more than one thinker who never called himself or his opinions by those appellations, and carefully guarded himself against being confounded with those who did, finds himself, sometimes to his displeasure, though generally by a tolerably correct instinct, classed with Positivists, and assailed as a Positivist."[1]

[1]This was expressed more recently by Kołakowski (1968, p.1) thus: "Not all, however, who according to historians or critics profess the positivist doctrine, would agree to be classified under this heading."

While a simplified version of positivistic science has come to be thought of as a philosophy within our everyday understanding of knowledge, most scientists acknowledge the metaphysical nature of such a position. Thus, most scientists have come, at least implicitly, to agree with Huxley's assertion, despite the claims of early positivists like Auguste Comte, that it "is impossible to live without metaphysic" (1937, p.252). To say that we should "let the data speak for themselves" contains the belief that *these* data are important, that they have something to say, and that it is worth hearing them out. However, most people's understanding of science and matters of methods or metaphysics tend to be on precisely such a "bird's eye" level, and the view of knowledge as being based on value-free, "objective" science is widespread (Aeschliman, 1983; Porter, 1996). "Decide on the basis of facts alone" became a mantra in modern times, and it remains so for many today, even among research psychologists. The continued dominance of positivism within psychology (Steinmetz, 2005), both as method and philosophy, shows that such thinking is widespread among social scientists who are aware of postmodern objections to positivistic epistemology. While various fields have gone through a wave (or waves) of intense empirical materialism, followed by periods of distancing from it, much of mainstream psychology appears to remain largely trapped in what Max Weber called the "iron cage of rationality," whereby not only is validity determined on the basis of quantification and empirical measurement, manipulation, and control, but also all other forms of knowledge are deemed secondary, if not irrelevant (Danziger, 1990; Valsiner, 2012). Put even more plainly and at the risk of oversimplification—we remain impressed by numbers. The bittersweet humor of the aphorism that "90% of statistics are made up on the spot" speaks not to the problem of statistics or quantification *per se*, but to the widespread, exaggerated, and even problematic nature of their persuasive power (Porter, 1996).

MAD AS A HATTER: THE "MADNESS" OF POSITIVISM AND OF POSTMODERN RESPONSES

> A flippant person has asked why we say, "As mad as a hatter." A more flippant person might answer that a hatter is mad because he has to measure the human head. (Chesterton, 1908/2015, p.7)

Extreme positivism has indeed led to various forms of "madness," but so too has the postmodern response to it. This is what led Laudan to call postmodern thinkers "the *new* crazies" (1996, p.3, italics added), a phrase which suggests the madness of both groups. As the general confusion sewn by postmodernism is well-known and widely discussed (e.g., Sokol and Bricmont, 1998; Valsiner, 2009; Hicks, 2011; Scruton, 2019), we will devote more time here to the madness of positivism. We will now briefly examine an illustrative example of the madness that can arise from the positivistic project.

Singer (2013), a moral philosopher and animal rights activist, made a telling utilitarian argument on the basis of positivistic *methods* turned positivistic *philosophy* for "effective altruism" (elsewhere called "outcome-based funding"). Effective altruism asserts that the quality of a charity can be determined on the basis of the number of quantifiable outcome units that result from each monetary unit invested in the given charity. Singer suggests that since it costs on average 40,000 United States dollars to raise and train a seeing-eye dog for a single blind American, but that the same amount of money could pay for operations that would allow between 400 and 2,000 people to regain their eyesight in the developing world (e.g., to treat glaucoma or cataracts), we should all redirect our donations from the first to the second kind of charity. Once the utilitarian ground has been laid, Singer points to the suffering, not of 2,000 sentient beings, but of literally billions of animals on the planet due to meat production, deforestation, and other forms of mistreatment. If we are counting, that number is certainly impressive and indeed heartbreaking, so perhaps that cause is an even better, more deserving recipient of our charitable donations. Singer does not stop there. Just as one blind person can be trumped by 2,000 blind people, and 2,000 blind people can be trumped by billions of suffering animals, so too can billions of suffering animals be trumped by the potential destruction of the entire planet along with all the creatures, big and small, living thereon. Thus, suggests Singer, perhaps we should be giving our charitable donations to the development of technologies that would shoot potentially Earth-ending asteroids out of the sky (or rather, shoot them to smithereens before they get that close). Rather than donating to afterschool programs for at-risk youth, suicide prevention hotlines, or cultural centers for the elderly, the quantification inherent in "effective altruism" would have us send checks to the space program. Here, decisions are made on the basis of mathematical calculation; responsibility lies in the tools, rather than in our own hands.

Postmodernism is currently one of the main voices in the social sciences that stands actively against the overly restrictive *Menschenbild* of positivism (Hicks, 2011). Much like the proliferation of new Protestant denominations after Catholic authority had been challenged, the notion of postmodernism can be thought of as an umbrella term that includes a large, and seemingly ever larger, number of different schools of thought (Steinmetz, 2005). Collectively, postmodernism is a set of inherently "critical, strategic and rhetorical practices [. . .used. . .] to destabilize other concepts such as presence, identity, historical progress, epistemic certainty, and the univocity of meaning" (Aylesworth, 2015). An attempt to define postmodernism, or even to explore it to any satisfactory degree, is well beyond the scope of the current piece. After all, "[t]hat postmodernism is indefinable is a truism" (Aylesworth, 2015). Postmodern thinkers have done much to shake the overly restrictive foundations of positivism, often by means of the intuitively unsatisfying nature of positivism, and they certainly have captured the imagination of many a social scientist and "layperson" alike (Sokol and Bricmont, 1998; Hicks, 2011; Scruton, 2019). Postmodernism has time and again illustrated how our lived experience in effect simply slips through the bars of positivism's methodological and epistemological iron cage. It has defended a wider scope of epistemological validity than that offered by positivistic reductionism alone (Steinmetz, 2005). However, postmodernism's primarily deconstructive nature remains inherently wedded to positivism, as its negation. Thus,

while postmodernism rightly challenges positivism's expansion of usefully reductionistic method into overly restrictive philosophy, it has been unable to offer satisfying methodological or philosophical alternatives (Sokol and Bricmont, 1998; Hicks, 2011; Scruton, 2019). "[P]ostpositivism is an intellectual failure. The arguments on its behalf are dubious and question-begging. Still worse, it has sustained virtually no positive program of research" (Laudan, 1996 p.5). Within psychology, "the post-modernism avalanche has been the replacement of inquiry by an epistemological battlefield" (Valsiner, 2009, p.2). However, the failure of postmodernism is not necessarily *sui generis*:

> what proved to be the undoing of postpositivism was not its departures from the positivist orthodoxy that preceded it. Rather, what has doomed postpositivism to amount to little more than a hiccup in the history of epistemology is the fact that it has carried to their natural conclusion several tendencies indigenous to positivism itself—tendencies that, once one sees their full spelling out, turn out to be wholly self-defeating. (Laudan, 1996, p.6)

Thus, postmodernism's "barbaric vagueness" can itself be understood as arising in large part from earlier attempts to "measure the head" (Tolman, 1992). It is therefore to early criticisms of positivism that we now turn to.

SAPIENTIA AS AN EARLY OBJECTION TO POSITIVISM

Challenges to the kinds of materialistic reductionism and quantification seen above are nothing new. For example, in contrast to the rationality of the Enlightenment, the likes of Vico and Herder asserted the fundamental importance of what Theodor Lipps later called empathetic *Einfühlung*, whereby we "feel into" the lives of qualitatively different others (Berlin, 1976). Similarly, as soon as psychology and sociology appeared at the university as independent disciplines at the turn of the 20th century, social scientists struggled to assert the fledgling fields' independence of philosophy on the basis of empirical research, while being simultaneously aware of the very serious limitations of reductionism (usually quantification and control), causal explanation, and prediction (Bruner, 1990). Even the founder of the first psychology laboratory in 1879, Wilhelm Wundt, became concerned with the overly positivistic turn the field was taking and thus attempted to promote a more historically oriented, interpretive cultural psychology (*Völkerpsychologie*) as a form of counterbalance (Bruner, 1990; Valsiner, 2012). Similar to Wundt, Max Weber thought that our ability to predict human behavior was improved by the social sciences, but that it was also limited. He believed the best we could achieve in combining scientific methods and human variability was "adequate causality," an approach that still very much describes the actual practices of statistical analysis in the social sciences today, even if hidden behind more strongly positivistic language (Tolman, 1992; Ringer, 2004). While he believed in the value of what he called the "ethic of responsibility" (*Verantwortungsethik*), whereby meaning is determined on the basis of causal relations in the empirical world, he was equally convinced of the need for the "ethic of conviction" (*Gesinnungsethik*) whereby the ends of human action, in the form of non-quantifiable values and meanings, must necessarily guide human life (Weber 1903–06/1975, p.192). Thus, according to Weber, the "social sciences, which are strictly empirical sciences, are the least fitted to presume to save the individual the difficulty of making a choice" (1949, p.19). Even the practicing scientist cannot escape living in a world that extends beyond the reductionistic horizon of positivism; "It can never be the task of an empirical science to provide binding norms and ideals from which directions for immediate practical activity can be derived" (Weber, 1949, p.52).

> Perhaps the nearest we can get to expressing it is to say this: that his mind [the follower of scientism] moves in a perfect but narrow circle. A small circle is quite as infinite as a large circle; but, though it is quite as infinite, it is not so large. (Chesterton, 1908/2015, p.8)

This short quotation wonderfully captures both the tremendously broad, and simultaneously limited, scope of positivism. The power of the scientific method for better understanding and manipulating the material world is undeniable. At the same time, in describing positivistic science as infinite but nevertheless "not so large" as *sapientia*, Chesterton uses the language of quantification to poetically evoke that which cannot be quantified, that which lies beyond even the infinite reach of science.

In the modern world, it has become increasingly difficult to speak about wisdom because we have indeed become trapped in what Max Weber called the "iron cage" of rationality. We have come to expect small, bit-sized slices of information to satisfy our search for knowledge, and to save us the difficulty of making discriminating judgments (Scruton, 2007). This can be clearly seen in the example of Singer's utilitarian assessment of charitable donations. Broadly speaking, this is a truly daunting problem. How can we consciously and conscientiously reflect on the strengths and weaknesses of positivistic epistemology when that has become the dominant epistemological language we have come to speak? Wittgenstein (1953) expressed something similar when he wrote: "The existence of the experimental method makes us think we have the means of solving the problems which trouble us; though problem and method pass one another by" (p.232). So as to break free of such positivistic language, advocates of *sapientia* have generally spoken in images or pictures rather than facts or data, in terms of qualities rather than quantities, in the language of poetry rather than prose, with the aid of judgment-provoking questions rather than unequivocal answers. In opposing the rigidity of positivism, thinkers such as G. K. Chesterton, C. S. Lewis, and L. Wittgenstein, generally wrote in a fantastical, poetic style, and often even in the language of fantasy or poetry themselves. To be clear, these thinkers were not against science, in the form of positivism-as-method. In fact, quite the opposite was generally the case. However, they were against *scientism*; the worship of a powerful method that, by default and design, constructively impoverishes our view of life and our *Menschenbild*.

> This is the paradox of imagination in science, that it has for its aim the impoverishment of imagination. By that outrageous phrase, I mean that the highest flight of scientific imagination is to weed out the proliferation of new ideas. In science, the grand view is a miserly view, and a rich model of the universe is one which is as poor as possible in hypotheses. (Bronowski, 1964, p.46)

The power of this impoverishment has been historically connected to Hume's separation of "is" from "ought," that is, the separation of inductive observations from normative positions (Brinkman, 2019). This is precisely the split that Weber and others (e.g., Popper) both acknowledged and found problematic. While some have argued the division between "matters of fact" and "matters of concern" to be largely artificial (Knorr-Cetina, 1991; Latour, 2004), both the power and pull of the scientific method is all but undeniable, and it echoes even in postmodernism (i.e., as its rebuttal). From the point of view of *sapientia*, it is not that scientific reductionism, causal explanations, or attempts at prediction are problematic in themselves. In fact, they are powerful and incredibly useful tools. Rather, the problem is that they have come to constrict our field of vision, leaving us with a myopic focus on the tools before us rather than a broader view of the wider world. Just as one should not forget the house for the hammer and the nails, one should not confuse the house with a home. In the words of C. K. Lewis, "second things suffer when put first" (in Aeschliman, 1983, p.33). This is similar to what Ortega y Gasset, 1948/1968, p.19) called the dehumanizing effect of "inhuman inversion" and what Polanyi (1958) called "moral inversion." If we understand science as a tool—not the only tool at our disposal, although a very important one—for the study of our lives and our world, we will be able to once again (re)focus on the important questions, rather than the questionnaires, that are driving our investigations.

Given our strong attachment to positivism in psychology, we may find ourselves still wondering at this point what exactly *sapientia* is, which is to say that we want a clear, positivistic definition. While we can confidently assert that the search for wisdom is roughly as old as humankind and that this search is found across cultures (Speer, 2005; Staudinger and Glück, 2011), singular definitions are inherently unsatisfying. Attempts to unilaterally define the notion belies the pull of positivism and the processes of increasing rationalization (Weber, 1949; Harrison, 2015). Some have even argued that "wisdom is the prototype of the class of psychological phenomena that by definition are unapproachable and unexplainable through scientific analysis" and that "to make wisdom transparent and to transform it into a subject matter of public knowledge and scientific debate is bound to change its basic foundation" (Baltes and Smith, 1990, p.89). This challenge aside, there have been considerable efforts by psychologists to conceptualize and operationalize wisdom so as to turn it into a measurable variable for empirical study (for overviews see Baltes and Smith, 2008; Staudinger and Glück, 2011; Ferrari and Weststrate, 2012; Bangen et al., 2013). These efforts are certainly laudable and have yielded valuable insights into the nature of our psychological lives. Nevertheless, to assume that this line of empirical research can capture the richness of wisdom is to misunderstand the issue at hand (Midgley, 1989; Maxwell, 2007). This is an example of Bronowski's "impoverishment of imagination" whereby we meaningfully and usefully make sense out of an otherwise essentially overwhelming phenomenon, forgetting in the process how much is necessarily lost in translation. While this approach can be valuable, it is also precisely what allows postmodernism to meaningfully object; in this case arguing that wisdom has been too narrowly and rigidly defined. However, in moving away from the ostensible clarity of empirically grounded conceptualizations and operationalizations, postmodernism turns quickly into a form of "anything goes" that renders the concept subjective to the point of meaninglessness. While positivism pushes figures and facts to the point of fault, postmodernism does so with the world of subjective feelings. By contrast, and at the risk of oversimplification, the approach of *sapientia* stresses neither figures nor feelings, but the elevating aspect of fantasy. It does not attempt to definitively solve the challenges before us (as do the positivists), nor does it attempt in essence to deny the existence of those challenges (as do the postmodernists). Rather, it recognizes the perennial relevance of the questions. Within clinical psychology, for example, while positivistic approaches would attempt to identify answers to questions of mental health, and postmodern approaches would illustrate the relative and subjective nature of both the questions and the answers, *sapientia* would remind us of the importance of reflecting again and again on the questions, e.g., what is mental health?

Thus, *sapientia* escapes clear, fixed definition. Like the haiku, it breaks free of representation, be it in numbers, words, linearity, circularity, and sequentiality, etc., while simultaneously avoiding the "anything goes" aspect of postmodernism. This understanding of *sapientia* is metaphorically explained by G. K. Chesterton (1908/2015 p.14) as follows: "The one created thing which we cannot look at is the one thing in the light of which we look at everything. Like the sun at noonday, mysticism explains everything else by the blaze of its own victorious invisibility." [. . .] "But the circle of the moon is as clear and unmistakable, as recurrent and inevitable, as the circle of Euclid on a blackboard. For the moon is utterly reasonable; and the moon is the mother of lunatics and has given to them all her name." Below we briefly review a few examples of how *sapientia* can reassert itself when we draw our attention back from a scientistic attachment to method, or what Gordon Allport called "methodolatry" (cited in Bruner, 1990, p.xi).

The Reinstatement of the Individual

The more science-as-method became science-as-philosophy, the less relevant became the individual to our overall intellectual pursuit in psychology. The reductionistic materialism of positivism has tended to reduce the attention paid to the individual by placing an increased focus on the aggregate. Danziger (1990) referred to this as the "triumph of the aggregate." Measures of central tendency are highlighted and outliers ignored or removed. This can allow us to see general principles beyond individual data points, and it is from this that science generally gains its power. While ideographic research has had a long and ongoing influence on the development of the sciences, including psychology, it is generally only considered to constitute scientific

knowledge once it has been extended beyond the individual (Salvatore and Valsiner, 2010). Numerous breakthroughs in psychology occurred in precisely this manner (e.g., memory research with patient "H.M."), as has been the case in other fields as well (e.g., studying individual planets or other cosmic bodies that are not easily replicable).

The strength of this epistemology-of-the-aggregate is also reflected in postmodernism in at least two interesting ways. On the one hand, as a reaction to positivism, postmodernism explicitly attempts to negate the epistemological certainty it affords. Thus, aggregate-level data are largely rejected, and the lion's share of attention within postmodern psychology is received by the individual case, especially the explicitly idiosyncratic case that cannot be readily linked into a larger collective or calculated into averages (such cases can, of course, also be on the group level, for example, by focusing on a particular collective). However, lest postmodernism fall into epistemic certainty on the individual level, following the exploration of a particular case, one often sees assertions of wider polyvalence, whereby the singular voice in question is but one among many such voices. In the words of Ernest E. Boesch, "a broom is a broom is a broom," by which he meant there are many different ways of seeing even a single, simple object (cited in Straub and Weidemann, 2007). On the other hand, despite such attempts to reject the modern epistemology-of-the-aggregate, postmodernism often sneaks such positivistic thinking in by the back door. For example, even when acknowledging the researcher's necessarily subjective position, postmodern psychology attempts to retain some sense of ("scientific") objectivity. This often takes the form of tool- or method-based distancing from the object of study, as such distance is understood—in line with the positivistic notion of objectivity (Porter, 1996; Daston and Galison, 2007)—to assure the validity of their truth claims about those same objects. For example, such notions as reaching the saturation point across interviews or setting a concrete number of interview subjects "required" to reach validity now seen in many journals, are in fact doing epistemology-of-the-aggregate without numbers (Holzkamp, 2013; Sousa, 2014; Huniche and Sørensen, 2019). Such methods are in effect different versions of letting the "data speaking for themselves"; the conclusions are being drawn from the objects of study via presumably "neutral" tools, and they are assumed to not be coming directly from the researcher, while they remain in many cases deeply method drive (Huniche and Sørensen, 2019). Thus, the ostensible openness and judgment-free position of the postmodern researcher, similar to the tool-based or method-based mediation of the natural sciences and positivistic psychology, distance the researcher from their research object, suggest the data to speak for themselves, and allow for claims of objectivity. This is psychology's version of what Whitehead (1920) called the bifurcation of nature, whereby the observer's tools allow them to objectively observe the observed. Thus, postmodernism tends to involve the explicit denial of the epistemology-of-the-aggregate, while simultaneously implicitly retaining it, although in a more subtle form.

Advocates of *sapientia* are not necessarily concerned with aggregate data, but by what Passmore (1978) called the "de-anthropomorphization" of human beings that can arise when the aggregate is valued above the individual. Historian Arnold Toynbee (1972/2015) was concerned by the "fanatical worship of collective human power," something that C. K. Lewis referred to as "that hideous strength" (Aeschliman, 1983, p.27). Incidentally, it is not surprising that psychology's maturity as a scientific and academic discipline at the start of the 20th century was marked by its close association with eugenics, perhaps the paradigmatic example of valuing the collective over the individual (Yakushko, 2019). A foreshadowing of this can be heard a century earlier in Comte's lament regarding "the perennial Western malady, the revolt of the individual against the species" (cited in Hayek, 1944/2007, p.70). Advocates of *sapientia* thus share postmodernism's assertion of the individual in abstraction from the collective, the promotion of the individual data point apart from the aggregate, in advocating for the outlier rather than the mean. However, *sapientia* differs from postmodernism by virtue of its simultaneous embrace of science *as method*, including the aggregate-level insights that come with it. What is more, *sapientia* asserts the value of the individual and the subjective as metaphysical values, a claim that postmodernism, in its general eschewal of epistemological certainty, would tend to avoid.

The Notion of Progress

Positivism is not only committed to the "infinite but narrow" circle of various forms of reductionism, but it is also deeply wedded to the notion of linear, unidirectional progress (Bury, 1920; Löwith, 1949). Positivists, including many psychologists today, believe (either explicitly or implicitly) that science marches progressively forward, and that with each new step we are that much closer to building a better, and perhaps even more perfect, world. The evolution of this particular understanding of linear, material progress arose over millennia, from the cyclical understanding of history of the Greeks and Romans to the Christian understanding of history that grafted otherworldliness (i.e., of salvation in the afterlife) onto Judaic historical linearity (Löwith, 1949). Positivism arose from this linear understanding of history, but rejects the notion of otherworldly hope, placing its telos in the material world. The hope of positivism lies in material progress and the (often implicit) belief in the perfectibility of the world. Within modern positivistic thinking, this understanding of progress has been further supported by numerous processes of conceptual reframing, such as the tendency to retrospectively (re)define "successes" (i.e., past positions that are supported today) as part of science, and "failures" (i.e., previously held positions that are rejected today) as unscientific (Sismondo, 1996). In this way, science appears to be always moving forward and on the right path.

While art history is certainly a serious academic discipline, few historians of art would argue that art has advanced in a linear manner over the centuries, and even fewer would profess a faith in the eternal, progressive march forward of the arts. Beauty has not gotten more beautiful, nor has our understanding of beauty been progressively improving. However, many scholars working in the natural sciences, such as chemistry or physics, would make what are at heart largely positivistic arguments regarding the evolution of their fields (even though these arguments would

be more nuanced today than pure, classic positivism). The scientism and methodolatry of psychology encourage a similar understanding of the field as progressing in a linear fashion. Empirical research in psychology, in the form of "neat little studies" (Bruner, 1990, p.xi) is expected to explicitly build, step-by-step, toward an ever-better understanding of human psychological processes. By default, and all things being equal, that which we know today is thought to be better than that which we knew yesterday, but inherently, an impoverishment of what we will know tomorrow.

With its focus on empirical research within "neat little studies" and on the "publish-or-parish" model of career development (and survival), contemporary psychology has witnessed an exponential growth of data collection. Within psychology, the rate of data collection has far outpaced the development of theory (Valsiner, 2014), reminding us of Percy Williams Bridgman's statement regarding mathematics: "As at present constructed, mathematics reminds one of the loquacious and not always coherent orator, who was said to be able to set his mouth going and go off and leave it" (cited in Sheen, 1934/2019, p.69). In addition to the numerous practical reasons for this development, one of the additional factors implicitly underlying it and implicitly justifying it is the positivistic belief in the forward march of progress on the basis of empirical data. The more data, the better. We have come to assume that the accumulation of facts equals the accumulation of knowledge. However, "Facts as facts do not make scientific knowledge" – "Experiments may abound, but there is no necessary increase in knowledge" (Sheen, 1934/2019, pp.60–61).

To the extent that postmodernism undercuts epistemic certainty, the notion of unilinear progress is an impossibility. In the face of multiple truths, not only is forward motion impossible to identify, but so too is a singular, "correct" path along which we may travel. Postmodernism therefore rejects the utopianism of positivism, and its understanding of polysemy often even challenges singular claims to progress. Like postmodernism, *sapientia* is not wedded to such a unidirectional notion of progress as is positivism. Within *sapientia*, the choices of today are not understood to be necessarily better than the ones of yesterday, and there is no necessary link between wisdom gained and the development of a better world. By seeing method as method, and not philosophy, *sapientia* promotes reflection on questions that are long-lasting if not eternal, even as particular puzzles about the material world might be solved. For example, even as we unravel many a mystery regarding particular issues of mental health (e.g., treating syphilitic dementia with penicillin), the question of what mental health is remains. Science is uniquely powerful for solving particular puzzles related to the material world, but it does not answer the fundamental questions of life. By shifting the telos back from the material world to broader metaphysical questions, claims of unilinear progress and utopian visions of tomorrow become more difficult to sustain. By the same logic, *sapientia* would not automatically subscribe to the assumed value of increasing empirical data collection within more and more "neat little studies." However, unlike postmodernism, *sapientia* asserts the value of the larger questions, not because they can be unilaterally answered as asserted by the positivist, but because they give life a simultaneously perennial and variable directionality that is generally undercut in postmodernism. This point will become clearer when we examine contradictions below.

Contradictions

The positivistic understanding of progress generally implies a rejection of contradictions. If two scientific claims contradict each other, something is amiss. Of course, in practice, scientists can come to contradictory conclusions for any number of reasons, and in fact, such tension lies at the heart of the scientific enterprise itself. Nevertheless, the power of science comes from its ability to make distinctions between competing theories and to explain or otherwise reconcile differences in the data. This is arguably broadly similar to the galvanizing and creative role of conflict seen in more specific schools of thought, such as conflict theory or Hegelian dualism, whereby social progress is made not on the basis of contradictions or conflicts themselves, by in the form of some sort of resolution (which is incidentally a nice example of how the postmodern search for conflict is deeply rooted to the positivistic search for resolution, Laudan, 1996). Within clinical and developmental psychology, this often appears in the form of various "crises," which lead to psychological growth (for an exploration of the "either-or" choice between objectivism and subjectivism in psychology, see Mos and Boodt, 1992).

That the need to resolve contradictions lies at the heart of the positivistic enterprise can be clearly seen in mainstream, quantitative psychology when experiments produce inconsistent or even contradictory findings. Researchers often work very hard, if not to definitively resolve them, then to explain them away by the addition of yet more variables; thereby kicking the proverbial can further down the road. An example of this can be seen in mainstream psychological research on power. Embedded within a larger positivistic research paradigm (including reductionistic conceptualizations and operationalizations, assumptions of variable control, the separation of variables, claims of causality, and replicability, etc.), high levels of power have been reported to make people lazy information processors, relying on previously held heuristics, and less flexible in the face of new information; however, other studies have found high levels of power to make people efficient information processors, relying less on previously held heuristic rules, and being more flexible in the face of new information (Guinote, 2015; Mazur, 2015). Such contradictory findings led to the formulation of the *situated theory of power* (Guinote, 2010), whereby variable, and even contradictory, information processing strategies are possible at high levels of power… depending on other variables (e.g., differences in motivation, goals, and situational factors). In light of such "variable buttressing," whereby new independent variables are added to hold up the presumed causal effects of other variables, it is not surprising that the main concept of interest itself slips further and further away. From variable buttressing emerge such claims as the following: "It may be less useful to seek a unified definition of power than to focus on systematic mapping of how the effects of power covary with the kind of power studied; that is, perhaps we are always consigned to study just one limited aspect of power at a time, but we can do so deliberately and explicitly, using multiple perspectives

and approaches in programmatic research" (Overbeck, 2010, p.32). Here we see a "three cup trick" at work, whereby the addition of yet more variables distracts us from the main point of interest; the definition of the construct (here "power") vanishes, while these definitional problems somehow do not exist for subcategories of the construct (here "various kinds of power") precisely because they lead to measurably different outcomes. Thus, within such positivistic research programs, the *definitional nature* of key concepts becomes their *functions* within larger causal chains (Mazur, 2015). This belief in the progressive value of adding rules to further explain previous rules, which were themselves expansions of previous rules, has been called the "additive fallacy" (Mazur, 2015). Interestingly, this process of smoothing out contradictions can be linked back to the drive for, and assumed value of, increased data collection and the linear progression of knowledge discussed earlier.

Postmodernism rejects this additive approach to knowledge and objects to this smoothing out of the rough edges of our lives in the name of singular explanations. Being reactionary in nature (i.e., "against" positivism), postmodernism actually highlights and promotes precisely such rough edges. In other words, postmodern thought is not so much an assertion of particular contradictions, but of contradiction itself. Within postmodernism, the contradiction is understood to be a tool to destabilize essentialized identities, the notion of historical progress, epistemic certainty, and the univocity of meaning. The postmodern voice is an "oppositional voice, a cry against the actual on behalf of the unknowable" (Scruton, 2019, p.15). Phenomenon and method are merged, and truth becomes indistinguishable from discourse. Sticking with the example of power, being "coextensive with the social body" (Foucault, 1980, p.142), power is understood in postmodernism as fluid and undefinable, but also ubiquitous. Opposition to power is inherently opposition to the social, particularly social stability, whatever that might mean in the given time and place.

Positivism is an attempt to remove contradictions, and postmodernism works to counteract those efforts. *Sapientia*, on the other hand, recognizes the insights gained by scientific discrimination (i.e., judging competing theories on the basis of data), while it also celebrates the contradictions of life—and it does so for their constructive, rather than deconstructive, value. In this tradition, contradictions, much like scientific tools, can be of epistemic value. Stated in the reverse, advocates of *sapientia* object to the reconciliation or resolution of contradictions seen in positivism (an objection shared with postmodernists), but they also object to the "contradiction-as-epistemic-uncertainty" found in postmodernism. *Sapientia* sees contradictions as a source of knowledge, as a way to break from the reductionistic rigidity of positivism. This particular celebration of contradictions has been called "bi-polar extremism" (Barron, 2004) within the Christian tradition, but it also finds expression in other religious, cultural, and historical contexts, such as the second century Mahayana Buddhist texts of Nāgārjuna (Garfield, 1995). The bipolar extremism of *sapientia* also finds expression in psychology, such as in Jungian archetypes (e.g., where each person contains light and shadow, male and female) or more recently in the notion of catalyzation within cultural psychology (Valsiner, 2014). Within *sapientia*, contradictions constitute paradoxes that inspire, not puzzles to be solved (as for the positivist) or perennial negations of truth at which we might throw up our hands in resignation or despair (as in postmodernism). The lover of *sapientia* "always cared more for truth than consistency. If he saw two truths that seemed to contradict each other, he would take the two truths and the contradiction along with them. His spiritual sight is stereoscopic, like his physical sight: he sees two different pictures at once and yet sees all the better for that" (Chesterton, 1908/2015, p.14). To stick with the metaphor of images, we can say that while positivism attempts to resolve the multistability of contradictions so that one truth wins out, and while postmodernism oscillates back and forth between the two *ad nauseam*, *sapientia* allows us to see both images at once (for a discussion of multistability see Mazur, 2019). With regard to power, perennial reflections on the nature of power lie at the heart of *sapientia*, and the questions that emerge are neither answered definitively away (as in positivism) nor deemed essentially subjective and unanswerable (as in postmodernism).

Wonder at the Ordinary

Our tendency to be enamored by positivistic reductionism, including quantification, has in effect distracted our attention from the actual object(s) of interest. This point can be seen in the quantification of psychological phenomena. Gustaw Ichheiser expressed this general challenge of positivistic psychology as follows: "the higher the adequacy of a psychological description, the stronger, paradoxically, the inevitable impression that 'nothing new' was really presented" (1943, p.207). The overquantification of psychological phenomena seen in modern psychology can cause a form of amnesia whereby we forget what we were interested in the first place:

> the psychologist may possess the knowledge about certain psychologically relevant facts as long as he is not acting in his capacity as a "psychologist." But, paradoxically enough, he forgets, ignores, or neglects those facts as soon as he transforms himself into a psychological expert and in this role performs some scientific research. (Ichheiser, 1943, p.206)

To be clear, the point is not that quantification cannot bring with it important insights regarding the nature of our world and our lives, including psychological phenomena. In fact, *sapientia* would not necessarily oppose the use of any tools at our disposal to better understand our lives. What is more, science—even strongly positivistic aspects of science—can certainly inspire wonder. This wonder cannot only inspire scientists in their work, but also it can often lead them to understand science to be, in fact, the ultimate source of wonder. Writing in the areas of physiology and psychology, Emil (Harleß, 1851, pp.20–21) captures this sentiment beautifully when he argues how the wonder of the natural world can remain hidden from what he calls "the unarmed eye" (*das unbewaffnete Auge*; without scientific tools like the microscope): "We are amazed by the beauty and the regularity with which the delicate plant cells are aligned with each other—and we trample with indifference the leaves of grass that are built out of them." However, the wonder such investigation can cause lies not in the numbers or

the methods or even the objects of investigation themselves, but in us (Polanyi, 1958; Wagoner et al., 2017). An awareness of the humanity of wonder is reflected in the stereotyped presentations of strictly positivistic minds in popular culture, such as Sherlock Holmes or the characters Mr. Spock and Data from Star Trek, characters whose excessively analytical minds are marked by a reduction in human emotion, including the sense of wonder. By way of contrast, Holmes' sidekick Dr. Watson is himself a scientist, but he is also deeply human, and he records the adventures of Sherlock Holmes precisely because of his sense of wonder at this uniquely overly analytical man. A similar contrast to the character of Spock is seen in the passionate character of Captain Kirk.

"And so we see that the poetry fades out of the problem and by the time the serious application of exact science begins we are left only with pointer readings" (Eddington, 1928/2012, p.252). It should not surprise us that the "pointer readings" of psychological research do not perfectly match the reality of our psychological lives. In fact, it should come as a great relief. As a response to positivism, *sapientia* not only redirects our attention to the ordinary, but reawakens our wonder at it. The ordinary not only constitutes the basic building blocks of our lives, but it is what makes our lives fantastic. As G. K. Chesterton put it, things that are common, like death or first love, are not necessarily commonplace; "Ordinary things are more valuable than extraordinary things; nay, they are more extraordinary" (1908/2015, p.26). Various schools of thought have emerged in psychology that constitute attempts to reinvigorate the study of those very topics that have been in decline during the ascent of positivistic thinking (i.e., various versions of "cultural psychology," not to be confused with cross-cultural psychology; Bruner, 1990; Lonner and Hayes, 2007; Valsiner, 2014). While such schools of thought generally lie outside the mainstream of psychology, they address issues that lie at the heart of the field, such as the study of mind rather than just the brain, the study of creativity rather than just causality, and the study of historically and socially based processes rather than presumed universality. These smaller branches of psychology thus share this fundamental similarity with early objections to positivism on the basis of *sapientia*; by redirecting our attention away from positivistic methods, they attempt to reawaken our wonder at that which has come to be seen as "ordinary" or that is made "normal" based on statistical normality. Here again, we hear echoes of postmodernism's objection to claims of normality (Scruton, 2019). However, unlike in postmodernism where the relativism of polyvalence curtails the depth and reach of wonder (as but one possible view), *sapientia* allows us to retain the object of wonder, precisely as an object of wonder.

DISCUSSION

Writing over a century ago, Georg Simmel argued that our increasing reliance on calculability and rationality (e.g., within the money-based economy) was having a significant impact on human psychology and social relations more broadly. This arose from:

> the growing preponderance of the category of quantity over that of quality, or more precisely the tendency to dissolve quality into quantity, to remove the elements more and more from quality, to grant them only specific forms of motion and to interpret everything that is specifically, individually, and qualitatively determined as the more or less, the bigger or smaller, the wider or narrower, the more or less frequent of those colorless elements and awarenesses that are only accessible to numerical determination. (Simmel, 1978/2004, p.278)

For the likes of Simmel and Weber, the problem was not so much with quantification itself, but with its imperial tendencies. Our reliance on calculability and rationality was changing the way we think and act. Something similar has happened in the field of psychology. Quantification has not only come to dominate as a methodological tool, but it has come to color our thinking about psychological phenomena more broadly. Despite the best efforts of postmodern thinkers, mainstream psychology retains a largely modern, positivistic epistemology.

Rather than rejecting the tools afforded us by positivistic thought, *sapientia* challenges us to recognize them as just that, tools. They are limited and necessarily unable to grasp the entirety of our psychological lives. Positivism is powerful as method; however, it is problematic as philosophy. To return to a metaphor used earlier, positivistic epistemology can help us identify the tools and the methods of their use by which we may build a more solid house, but it cannot make it a home nor even help us to better understand what that might require. In other words, positivism cannot fully capture the psychosocial meaning-making processes that should constitute the core focus of psychology (Bruner, 1990). In the words of Gustaw Ichheiser:

> We should not expect and demand that everything should be "proved." To say it once more, social scientists should, in my opinion, not aspire to be as "scientific" and "exact" as physicists or mathematicians, but should *cheerfully* accept the fact that what they are doing belongs to the twilight zone between science and literature. (cited in Rudmin et al., 1987, p.171, italics added)

It has been argued that we can indeed *cheerfully* accept this state of affairs, and that in doing so, we will be able to once again see aspects of our psychosocial lives that have become obscured by the dominant positivistic epistemology, such as the importance of the individual and of subjective experience, the notion of *personal* progress in a complex world, the epistemic power of contradictions, and a childlike wonder at our world. Positivistic epistemology in psychology is certainly powerful, but it is also of limited use; however, the way to approach those shortcomings lies not in postmodernism. Rather, we should more fully recognize that as psychologists, we "are still drawing rich sustenance from our more distant, pre-positivist past" (Bruner, 1990, p.x).

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Aeschliman, M. D. (1983). *The Restitution of Man: C. S. Lewis and the Case Against Scientism*. Grand Rapids, MI: William B. Eerdmans Publishing Company.

Aylesworth, G. (2015). "Postmodernism," in *The Stanford Encyclopedia of Philosophy (Spring 2015 Edn)*, ed. E. N. Zalta (Stanford, CA: Stanford University).

Baker, W. J. (1992). "Positivism versus people: what should psychology be about?," in *Positivism in Psychology. Historical and Contemporary Problems*, ed. C. W. Tolman (New York, NY: Springer), 9–16. doi: 10.1007/978-1-4612-4402-8_2

Baltes, P. B., and Smith, J. (1990). "Toward a psychology of wisdom and its ontogenesis," in *Wisdom: Its Nature, Origins, and Development*, ed. R. J. Sternberg (Cambridge: Cambridge University Press), 87–120. doi: 10.1017/cbo9781139173704.006

Baltes, P. B., and Smith, J. (2008). The fascination of wisdom. Its nature, ontogeny, and function. *Perspect. Psychol. Sci.* 3, 56–64. doi: 10.1111/j.1745-6916.2008.00062.x

Bangen, K. J., Meeks, T. W., and Jeste, D. V. (2013). Defining and assessing wisdom: a review of the literature. *Am. J. Geriatr. Psychiatry* 21, 1254–1266. doi: 10.1016/j.jagp.2012.11.020

Barron, R. (2004). *Bridging the Great Divide*. Oxford: Rowman & Littlefield.

Berlin, I. (1976). *Vico & Herder: Two Studies in the History of Ideas*. New York, NY: The Viking Press.

Brinkman, S. (2019). "Normativity in psychology and the social sciences: questions of universality," in *Social Philosophy of Science for The Social Sciences*, ed. J. Valsiner (Cham: Springer). doi: 10.1007/978-3-030-33099-6_11

Bronowski, J. (1964). *The Identity of Man*. Garden City, NY: American Museum Science Books.

Bruner, J. (1990). *Acts of Meaning*. Cambridge, MA: Harvard University Press.

Bury, J. B. (1920). *The Idea of Progress: An Inquiry into its Origin and Growth*. London: MacMillian & Co.

Caldwell, B. (1994). *Beyond Positivism*. New York, NY: Taylor & Francis.

Chesterton, G. K. (1908/2015). *Orthodoxy*. Middletown, DE: Snowball Classics.

Cornejo, C. (2017). "From fantasy to imagination. A cultural history and a moral for psychology," in *The Psychology of Imagination*, eds B. Wagoner, I. B. de Luna, and S. H. Awad (Charlotte, NC: Information Age Publishing), 3–46.

Danziger, K. (1990). *Constructing the Subject: Historical Origins of Psychological Research*. Cambridge: Cambridge University Press.

Daston, L., and Galison, P. (2007). *Objectivity*. Brooklyn, NY: Zone Books.

Eddington, A. (1928/2012). *The Nature of the Physical World*. Cambridge: Cambridge University Press.

Elliot, A. (2005). "Psychoanalysis as critique: psychoanalysis and the theory of the subject," in *The Politics of Method in the Human Sciences. Positivism and its Epistemological Others*, ed. G. Steinmetz (Durham, NC: Duke University Press), 427–450. doi: 10.1215/9780822386889-014

Feichtigner, J., Fillafer, F. L., and Surman, J. (eds) (2018). *The Worlds of Positivism: A Global Intellectual History*. Cham: Palgrave Macmillian.

Ferrari, M., and Weststrate, N. (eds) (2012). *The Scientific Study of Personal Wisdom: from Contemplative Traditions to Neuroscience*. New York, NY: Springer. doi: 10.1007/978-90-481-9231-1_15

Foucault, M. (1980). *Power/Knowledge: Selected Interviews and Other Writings*. New York, NY: Pantheon Books.

Frontiers in Psychology (n.d.). This Statement is Take From the Call for Papers for the Special Edition of Frontiers in Psychology Entitled "From 'Modern' to 'Postmodern' Psychology: is There a Way Past? Available online at: https://www.frontiersin.org/research-topics/11572/from-modern-to-postmodern-psychology-is-there-a-way-past (accessed October 27, 2020).

Garfield, J. L. (1995). *The Fundamental Wisdom of the Middle Way. Nāgārjuna's Mūlamadhyamakakārikā*. New York, NY: Oxford University Press.

Gozli, D. G., and Deng, W. (2018). Building blocks of psychology: on remaking the unkept promises of early schools. *Integrat. Psychol. Behav. Sci.* 52, 1–24. doi: 10.1007/s12124-017-9405-7

Guinote, A. (2010). "The situated focus theory of power," in *The Social Psychology of Power*, eds A. Guinote and T. K. Vescio (New York, NY: Guilford Publications), 141–173.

Guinote, A. (2015). "Social cognition of power," in *APA Handbook of Personality and Social Psychology*, Vol. 1, eds M. Mikulincer and P. R. Shaver (Washington, DC: American Psychological Association), 547–569. doi: 10.1037/14341-017

Harleß, E. (1851). *Populäre Vorlesungen aus dem Gebiet der Physiologie und Psychologie*. Braunschweig: Friedrich Vieweg und Sohn.

Harrison, P. (2015). *The Territories of Science and Religion*. Chicago: University of Chicago Press.

Hayek, F. A. (2007/1944). *The Road to Serfdom*. Chicago, IL: University of Chicago Press.

Hicks, S. R. C. (2011). *Explaining Postmodernism: Skepticism and Socialism from Rousseau to Foucault (expanded edition)*. Roscoe, IL: Ockham's Razor Publishing.

Holzkamp, K. (2013). "Missing the point: variable psychology's blindness to the problem's inherent coherences," in *Psychology from the Standpoint of the Subject: Selected Writings of Klaus Holzkamp*, eds E. Schraube and U. Osterkamp (London: Palgrave Macmillan), 60–74. doi: 10.1057/9781137296436_5

Huniche, L., and Sørensen, E. (2019). Phenomenon-driven research and systematic research assembling: methodological conceptualizations for psychology's epistemic projects. *Theory Psychol.* 29, 539–558. doi: 10.1177/0959354319862048

Hunter, M. (2000). *Robert Boyle (1627-91): Scrupulosity and Science*. Woodbridge: The Boydell Press.

Huxley, A. (1937). *Ends and Means*. London: Chatto & Windus.

Ichheiser, G. (1943). Why psychologists tend to overlook certain "obvious" facts. *Philos. Sci.* 10, 204–207. doi: 10.1086/286811

Knorr-Cetina, K. (1991). Epistemic cultures: forms of reason in science. *Hist. Politi. Econ.* 23, 105–122. doi: 10.1215/00182702-23-1-105

Kołakowski, L. (1968). *The Alienation of Reason: A History of Positivist Thought*. Garden City, NY: Doubleday.

Latour, B. (2004). Why has critique run out of steam? From matters of fact to matters of concern. *Crit. Inq.* 30, 225–248. doi: 10.2307/1344358

Laudan, L. (1996). *Beyond Positivism and Relativism: Theory, Method and Evidence*. Boulder, CO: Westview.

Lonner, W. J., and Hayes, S. A. (2007). *Discovering Cultural Psychology: A Profile and Selected Readings of Ernst E. Boesch*. Charlotte, NC: Information Age Publishing.

Löwith, K. (1949). *Meaning in History: The Theological Implications of the Philosophy of History*. Chicago, IL: University of Chicago Press.

Maxwell, N. (2007). *From Knowledge to Wisdom: A Revolution for Science and the Humanities*, 2nd Edn, London: Pentire.

Maxwell, N. (2017). *Karl Popper, Science and Enlightenment*. London: UCL Press. doi: 10.2307/j.ctt1vxm8p6

Mazur, L. B. (2015). Defining power in social psychology. *Orbis Idea.* 2, 101–114.

Mazur, L. B. (2017). "Gaps in human knowledge: highlighting the whole beyond our conceptual reach," in *The Psychology of Imagination: History, Theory and New Research Horizons*, eds B. Wagoner, I. Bresco, and S. H. Awad (Charlotte, NC: Information Age Publishing), 239–252.

Mazur, L. B. (2019). The semiotic multistability of time: literature as a window onto a classic challenge in psychology. *Theory Psychol.* 29, 358–376. doi: 10.1177/0959354319848496

Mazur, L. B., and Watzlawik, M. (2016). Debates about the scientific status of psychology: looking at the bright side. *Integr. Psychol. Behav. Sci.* 50, 555–567. doi: 10.1007/s12124-016-9352-8

Midgley, M. (1989). *Wisdom, Information, and Wonder: What is Knowledge for?* London: Routledge.

Mill, J. S. (1865/2005). *Auguste Comte and Positivism. Project Gutenberg*. Available online at: (accessed October 27, 2020).

Mos, L. P., and Boodt, C. P. (1992). "Mediating between mentalism and objectivism: the conventionality of language," in *Positivism in Psychology: Historical and Contemporary Problems*, ed. C. W. Tolman (New York, NY: Springer), 185–216. doi: 10.1007/978-1-4612-4402-8_13

Ortega y Gasset, J. (1948/1968). *The Dehumanization of Art and Other Essays on Art, Culture, and Literature*. Princeton, NJ: Princeton University Press.

Overbeck, J. R. (2010). "Concepts and historical perspectives on power," in *The Social Psychology of Power*, eds A. Guinote and T. K. Vescio (New York, NY: Guilford Publications), 19–45.

Passmore, J. (1978). *Science and its Critics*. London: Duckworth Press.

Polanyi, M. (1958). *Personal Knowledge*. Chicago, IL: University of Chicago Press.

Porter, T. M. (1996). *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Ringer, F. (2004). *Max Weber: An Intellectual Biography*. Chicago, IL: University of Chicago Press.

Rudmin, F., Trimpop, R. M., Kryl, I.-P., and Boski, P. (1987). Gustav Ichheiser in the history of social psychology: an early phenomenology of social attribution. *Br. J. Soc. Psychol.* 26, 165–180. doi: 10.1111/j.2044-8309.1987.tb00777.x

Salvatore, S., and Valsiner, J. (2010). Between the general and the unique: overcoming the nomothetic versus ideographic opposition. *Theory Psychol.* 20, 817–833. doi: 10.1177/0959354310381156

Schwarz, K. A., and Pfister, R. (2016). Scientific psychology in the 18th century: a historical rediscovery. *Perspect. Psychol. Sci.* 11, 399–407. doi: 10.1177/1745691616635601

Scriven, M. (1966). "The contribution of philosophy of the social sciences to educational development," in *Philosophy and Educational Development*, ed. G. Barnett (Boston, MA: Houghton Mifflin Company), 47–68.

Scruton, R. (2007). *Culture Counts: Faith and Feeling in a World Besieged*. New York, NY: Brief Encounters.

Scruton, R. (2019). *Fools, Frauds, and Firebrands: Thinkers of the New Left*. New York, NY: Bloomsbury.

Sheen, F. J. (1934/2019). *The Philosophy of Science*. Providence, RI: Cluny Media.

Simmel, G. (1978/2004). *The Philosophy of Money*, 3rd Edn, New York, NY: Routledge.

Singer, P. (2013). *The Why and How of Effective Altruism. Ted Talk*. Available online at: https://www.ted.com/talks/peter_singer_the_why_and_how_of_effective_altruism?language=en (accessed October 27, 2020).

Sismondo, S. (1996). *Science without Myth: On Constructions, Reality, and Social Knowledge*. Albany, NY: State University of New York Press.

Smythe, W. E. (1992). "Positivism and the prospects for cognitive science," in *Positivism in Psychology: Historical and Contemporary Problems*, ed. C. W. Tolman (New York, NY: Springer), 103–118. doi: 10.1007/978-1-4612-4402-8_9

Sokol, A., and Bricmont, J. (1998). *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science*. New York, NY: Picador.

Sousa, D. (2014). Validation in qualitative research: general aspects and specificities of the descriptive phenomenological method. *Q. Res. Psychol.* 11, 211–227. doi: 10.1080/14780887.2013.853855

Speer, A. (2005). "Weisheit," in *Historisches Wörterbuch der Philosophie*, eds J. Ritter, K. Gründer, and G. Gabriel (Basel: Schwabe).

Staudinger, U. M., and Glück, J. (2011). Psychological wisdom research: commonalities and differences in a growing field. *Annu. Rev. Psychol.* 62, 215–241. doi: 10.1146/annurev.psych.121208.131659

Steinmetz, G. (ed.) (2005). *The Politics of Method in the Human Sciences: Positivism and its Epistemological Others*. Durham, NC: Duke University Press.

Straub, J., and Weidemann, A. (2007). "Introduction to ideas: experience, method, and dynamic self: prefatory comments on Ernest E. Boesch's contributions to cultural psychology," in *Discovering Cultural Psychology: A Profile and Selected Readings of Ernst E. Boesch*, eds W. J. Lonner and S. A. Hayes (Charlotte, NC: Information Age Publishing), 39–58.

Tolman, C. W. (ed.) (1992). *Positivism in Psychology: Historical and Contemporary Problems*. New York, NY: Springer. doi: 10.1007/978-1-4612-4402-8

Toynbee, A. (1972/2015). "Introduction ," in *The Gods of Revolution*, ed. C. Dawson (New York, NY: New York University Press), xxi-xxiiii, 21–23.

Valsiner, J. (2009). Integrating psychology within the globalizing world: a requiem to the pos-modern experiment with Wissenschaft. *Integr. Psychol. Behav. Sci.* 43, 1–21. doi: 10.1007/s12124-009-9087-x

Valsiner, J. (2012). *A Guided Science: History of Psychology in the Mirror of its Making*. New Brunswick, NJ: Transaction Publishers.

Valsiner, J. (2014). *An Invitation to Cultural Psychology*. Washington, DC: Sage. doi: 10.4135/9781473905986

Wagoner, B., Luna, I. B., and Awad, S. H. (eds) (2017). *The Psychology of Imagination*. Charlotte, NC: Information Age Publishing.

Weber, M. (1949). "The meaning of ethical neutrality," in *Max Weber on the Methodology of the Social Sciences*, eds E. A. Shils and H. Finch (Glencoe, IL: The Free Press), 1–47. doi: 10.4324/9781315124445-1

Whitehead, A. N. (1920). *The Concept of Nature*. Cambridge: Cambridge University Press.

Wittgenstein, L. (1953). *Philosophical Investigations*, eds G. E. M. Anscombe and R. Rhees (Oxford: Blackwell).

Yakushko, O. (2019). Eugenics and its evolution in the history of western psychology: a critical archival review. *Psychother. Polit. Intern.* 17, 1–13. doi: 10.1002/ppi.1495

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 12 January 2021
doi: 10.3389/fpsyg.2020.612805

The Practice of Experimental Psychology: An Inevitably Postmodern Endeavor

Roland Mayrhofer, Christof Kuhbandner and Corinna Lindner*

Department of Psychology, University of Regensburg, Regensburg, Germany

The aim of psychology is to understand the human mind and behavior. In contemporary psychology, the method of choice to accomplish this incredibly complex endeavor is the experiment. This dominance has shaped the whole discipline from the self-concept as an empirical science and its very epistemological and theoretical foundations, via research practice and the scientific discourse to teaching. Experimental psychology is grounded in the scientific method and positivism, and these principles, which are characteristic for modern thinking, are still upheld. Despite this apparently stalwart adherence to modern principles, experimental psychology exhibits a number of aspects which can best be described as facets of postmodern thinking although they are hardly acknowledged as such. Many psychologists take pride in being "real natural scientists" because they conduct experiments, but it is particularly difficult for psychologists to evade certain elements of postmodern thinking in view of the specific nature of their subject matter. Postmodernism as a philosophy emerged in the 20th century as a response to the perceived inadequacy of the modern approach and as a means to understand the complexities, ambiguities, and contradictions of the times. Therefore, postmodernism offers both valuable insights into the very nature of experimental psychology and fruitful ideas on improving experimental practice to better reflect the complexities and ambiguities of human mind and behavior. Analyzing experimental psychology along postmodern lines begins by discussing the implications of transferring the scientific method from fields with rather narrowly defined phenomena—the natural sciences—to a much broader and more heterogeneous class of complex phenomena, namely the human mind and behavior. This ostensibly modern experimental approach is, however, *per se* riddled with postmodern elements: (re-)creating phenomena in an experimental setting, including the hermeneutic processes of generating hypotheses and interpreting results, is no carbon copy of "reality" but rather an active construction which reflects irrevocably the pre-existing ideas of the investigator. These aspects, analyzed by using postmodern concepts like hyperreality and simulacra, did not seep in gradually but have been present since the very inception of experimental psychology, and they are necessarily inherent in its philosophy of science. We illustrate this theoretical analysis with the help of two examples, namely experiments on free will and visual

OPEN ACCESS

Edited by:
Peter Adriaan Edelsbrunner, ETH Zürich, Switzerland

Reviewed by:
Peter Holtz, Leibniz-Institut für Wissensmedien (IWM), Germany
Tom Dickins, Middlesex University, United Kingdom

****Correspondence:***
Roland Mayrhofer
roland.mayrhofer@ur.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *30 September 2020*
Accepted: *26 November 2020*
Published: *12 January 2021*

Citation:
Mayrhofer R, Kuhbandner C and Lindner C (2021) The Practice of Experimental Psychology: An Inevitably Postmodern Endeavor. Front. Psychol. 11:612805. doi: 10.3389/fpsyg.2020.612805

working memory. The postmodern perspective reveals some pitfalls in the practice of experimental psychology. Furthermore, we suggest that accepting the inherently fuzzy nature of theoretical constructs in psychology and thinking more along postmodern lines would actually clarify many theoretical problems in experimental psychology.

Keywords: postmodernism, experimental psychology, experiment, methodology, philosophy of science

INTRODUCTION

Postmodernism is, in essence, an attempt to achieve greater clarity in our perception, thinking, and behavior by scrutinizing their larger contexts and preconditions, based on the inextricably intertwined levels of both the individual and the society. Psychology also studies the human mind and behavior, which indicates that psychology should dovetail with postmodern approaches. In the 1990s and early 2000s, several attempts were made to introduce postmodern thought as potentially very fruitful ideas into general academic psychology (Jager, 1991; Kvale, 1992; Holzman and Morss, 2000; Holzman, 2006). However, overall they were met with little response.

Postmodern thoughts have been taken up by several fringe areas of academic psychology, e.g., psychoanalysis (Leffert, 2007; Jiménez, 2015; but see Holt, 2005), some forms of therapy and counseling (Ramey and Grubb, 2009; Hansen, 2015), humanistic (Krippner, 2001), feminist and gender (Hare-Mustin and Marecek, 1988; Sinacore and Enns, 2005), or cultural psychology (Gemignani and Peña, 2007).

However, there is resistance against suggestions to incorporate postmodern ideas into the methodology and the self-perception of psychology as academic—and scientific!—discipline. In fact, postmodern approaches are often rejected vehemently, sometimes even very vocally. For instance, Gergen (2001) argued that the "core tenets" of postmodernism are not at odds with those of scientific psychology but rather that they can enrich the discipline by opening up new possibilities. His suggestions were met with reservation and were even outright rejected on the following grounds: postmodernism, "like anthrax of the intellect, if *allowed* [our italics] into mainstream psychology, [. . .] will poison the field" (Locke, 2002, 458), that it "wishes to return psychology to a prescientific subset of philosophy" (Kruger, 2002, 456), and that psychology "needs fewer theoretical and philosophical orientations, not more" (Hofmann, 2002, 462; see also Gergen's, 2001, replies to the less biased and more informed commentaries on his article).

In the following years, and continuing the so-called science wars of the 1990s (Segerstråle, 2000), several other attacks were launched against a perceived rise or even dominance of postmodern thought in psychology. Held (2007; see also the rebuttal by Martin and Sugarman, 2009) argued that anything postmodern would undermine rationality and destroy academic psychology. Similarly, postmodernism was identified—together with "radical environmentalism" and "pseudoscience" among other things—as a "key threat to scientific psychology" (Lilienfeld, 2010, 282), or as "inimical to progress in the psychology of science" (Capaldi and Proctor, 2013, 331). The following advice was given to psychologists: "We [psychologists] should also push back against the pernicious creep of these untested concepts into our field" (Tarescavage, 2020, 4). Furthermore, the term "postmodern" is even employed as an all-purpose invective in a popular scientific book by psychologist Steven Pinker (2018).

Therefore, it seems that science and experimental psychology on the one hand and postmodern thinking on the other are irreconcilable opposites. However, following Gergen (2001) and Holtz (2020), we argue that this dichotomy is only superficial because postmodernism is often misunderstood. A closer look reveals that experimental psychology contains many postmodern elements. Even more, there is reason to assume that a postmodern perspective may be beneficial for academic psychology: First, the practice of experimental psychology would be improved by integrating postmodern thinking because it reveals a side of the human psyche for which experimental psychology is mostly blind. Second, the postmodern perspective can tell us much about the epistemological and social background of experimental psychology and how this affects our understanding of the human psyche.

A POSTMODERN PERSPECTIVE ON EXPERIMENTAL PSYCHOLOGY

Experimental Psychology and the Modern Scientific Worldview

It lies within the nature of humans to try to find out more about themselves and their world, but the so-called Scientific Revolution of the early modern period marks the beginning of a new era in this search for knowledge. The Scientific Revolution, which has led to impressive achievements in the natural sciences and the explanation of the physical world (e.g., Olby et al., 1991; Henry, 1997; Cohen, 2015; Osterlind, 2019), is based on the following principle: to "measure what can be measured and make measurable what cannot be measured." This famous appeal—falsely attributed to Galileo Galilei but actually from the 19th century (Kleinert, 2009)—illustrates the two fundamental principles of modern science: First, the concept of "measurement" encompasses the idea that phenomena can be quantified, i.e., expressed numerically. Second, the concept of "causal connections" pertains to the idea that consistent, non-random relationships can be established between measurable phenomena. Quantification allows that relationships between phenomena can be expressed, calculated, and predicted in precise mathematical and numerical terms.

However, there are two important issues to be aware of. First, while it is not difficult to measure "evident" aspects, such as mass

and distance, more complex phenomena cannot be measured easily. In such cases, it is therefore necessary to find ways of making these "elusive" phenomena measurable. This can often only be achieved by reducing complex phenomena to their simpler—and measurable!—elements. For instance, in order to measure memory ability precisely, possible effects of individual preexisting knowledge which introduce random variance and thus impreciseness have to be eliminated. Indeed, due to this reason, in many memory experiments, meaningless syllables are used as study material.

Second, it is not difficult to scientifically prove a causal relationship between a factor and an outcome if the relationship is simple, that is, if there is only one single factor directly influencing the outcome. In such a case, showing that a manipulation of the factor causes a change in the outcome is clear evidence for a causal relationship because there are no other factors which may influence the outcome as well. However, in situations where many factors influence an outcome in a complex, interactive way, proving a causal relationship is much more difficult. To prove the causal effect of one factor in such a situation the effects of all other factors—called confounding factors from the perspective of the factor of interest—have to be eliminated so that a change in the outcome can be truly attributed to a causal effect of the factor of interest. However, this has an important implication: The investigator has to divide the factors present in a given situation into interesting versus non-interesting factors with respect to the current context of the experiment. Consequently, while experiments reveal something about local causal relationships, they do not necessarily provide hints about the net effect of all causal factors present in the given situation.

The adoption of the principles of modern science has also changed psychology. Although the beginnings of psychology—as the study of the *psyche*—date back to antiquity, psychology as an academic discipline was established in the mid to late 19th century. This enterprise was also inspired by the success of the natural sciences, and psychology was explicitly modeled after this example by Wilhelm Wundt—the "father of experimental psychology"—although he emphasized the close ties to the humanities as well. The experiment quickly became the method of choice. There were other, more hermeneutic approaches during this formative phase of modern psychology, such as psychoanalysis or introspection according to the Würzburg School, but their impact on academic psychology was limited. Behaviorism emerged as a direct reaction against these perceived unscientific approaches, and its proponents emphasized the scientific character of their "new philosophy of psychology." It is crucial to note that in doing so they also emphasized the importance of the experiment and the necessity of quantifying directly observable behavior in psychological research. Behaviorism quickly became a very influential paradigm which shaped academic psychology. Gestalt psychologists, whose worldview is radically different from behaviorism, also relied on experiments in their research. Cognitive psychology, which followed, complemented, and partly superseded behaviorism, relies heavily on the experiment as a means to gain insight into mental processes, although other methods such as modeling are employed as well. Interestingly, there is a fundamental difference between psychoanalysis and humanistic psychology, which do not rely on the experiment, and the other above-mentioned approaches as the former focus on the psychic functioning of individuals, whereas the latter focus more on global laws of psychic functioning across individuals. This is reflected in the fact that psychological laws in experimental psychology are established on the arithmetic means across examined participants—a difference we will elaborate on later in more detail. Today, psychology is the *scientific*—in the sense of empirical-quantitative—study of the human mind and behavior, and the experiment is often considered the gold standard in psychological research (e.g., Mandler, 2007; Goodwin, 2015; Leahey, 2017).

The experiment is closely associated with the so-called scientific method (Haig, 2014; Nola and Sankey, 2014) and the epistemological tenets philosophy of positivism—in the sense as Martin (2003); Michell (2003), and Teo (2018) explain—which sometimes exhibit characteristics of naïve empiricism. Roughly speaking, the former consists of observing, formulating hypotheses, and testing these hypotheses in experiments. The latter postulates that knowledge is based on sensory experience, that it is testable, independent of the investigator and therefore objective as it accurately depicts the world as it is. This means that in principle all of reality can not only be measured but eventually be entirely explained by science. This worldview is attacked by postmodern thinkers who contend that the world is far more complex and that the modern scientific approach cannot explain all of reality and its phenomena.

The Postmodern Worldview

Postmodern thinking (e.g., Bertens, 1995; Sim, 2011; Aylesworth, 2015) has gained momentum since the 1980s, and although neither the term "postmodernism" nor associated approaches can be defined in a unanimous or precise way, they are characterized by several intertwined concepts, attitudes, and aims. The most basic trait is a general skepticism and the willingness to question literally everything from the ground up—even going so far as to question not only the foundation of any idea, but also the question itself. This includes the own context, the chosen premises, thinking, and the use of language. Postmodernism therefore has a lot in common with science's curiosity to understand the world: the skeptical attitude paired with the desire to discover how things really are.

Postmodern investigations often start by looking at the language and the broader context of certain phenomena due to the fact that language is the medium in which many of our mental activities—which subsequently influence our behavior—take place. Thus, the way we talk reveals something about how and why we think and act. Additionally, we communicate about phenomena using language, which in turn means that this discourse influences the way we think about or see those phenomena. Moreover, this discourse is embedded in a larger social and historical context, which also reflects back on the use of language and therefore on our perception and interpretation of certain phenomena.

Generally speaking, postmodern investigations aim at detecting and explaining how the individual is affected by societal influences and their underlying, often hidden ideas, structures, or mechanisms. As these influences are often fuzzy, contradictory, and dependent on their context, the individual is subject to a multitude of different causalities, and this already complex interplay is further complicated by the personal history, motivations, aims, or ways of thinking of the individual. Postmodernism attempts to understand all of this complexity as it is in its entirety.

The postmodern approaches have revealed three major general tendencies which characterize the contemporary world: First, societies and the human experience since the 20th century have displayed less coherence and conversely a greater diversity than the centuries before in virtually all areas, e.g., worldviews, modes of thinking, societal structures, or individual behavior. Second, this observation leads postmodern thinkers to the conclusion that the grand narratives which dominated the preceding centuries and shaped whole societies by providing frames of references have lost—at least partially—their supremacy and validity. Examples are religious dogmas, nationalism, industrialization, the notion of linear progress—and modern science because it works according to certain fundamental principles. Third, the fact that different but equally valid perspectives, especially on social phenomena or even whole worldviews, are possible and can coexist obviously affects the concepts of "truth," "reality," and "reason" in such a way that these concepts lose their immutable, absolute, and universal or global character, simply because they are expressions and reflections of a certain era, society, or worldview.

At this point, however, it is necessary to clarify a common misconception: Interpreting truth, reality, or reason as relative, subjective, and context-dependent—as opposed to absolute, objective, and context-independent—does naturally neither mean that anything can be arbitrarily labeled as true, real, or reasonable, nor, vice versa, that something cannot be true, real, or reasonable. For example, the often-quoted assumption that postmodernism apparently even denies the existence of gravity or its effects as everything can be interpreted arbitrarily or states that we cannot elucidate these phenomena with adequate accuracy because everything is open to any interpretation (Sokal, 1996), completely misses the point.

First, postmodernism is usually not concerned with the laws of physics and the inanimate world as such but rather focuses on the world of human experience. However, the phenomenon itself, e.g., gravity, is not the same as our scientific knowledge of phenomena—our chosen areas of research, methodological paradigms, data, theories, and explanations—or our perception of phenomena, which are both the results of human activities. Therefore, the social context influences our scientific knowledge, and in that sense scientific knowledge is a social construction (Hodge, 1999).

Second, phenomena from human experience, although probably more dependent on the social context than physical phenomena, cannot be interpreted arbitrarily either. The individual context—such as the personal history, motivations, aims, or worldviews—determines whether a certain behavior makes sense for a certain individual in a certain situation. As there are almost unlimited possible backgrounds, this might seem completely random or arbitrary from an overall perspective. But from the perspective of an individual the phenomenon in question may be explained entirely by a theory for a specific—and not universal—context.

As described above, the postmodern meta-perspective directly deals with human experience and is therefore especially relevant for psychology. Moreover, any discipline—including the knowledge it generates—will certainly benefit from understanding its own (social) mechanisms and implications. We will show below that postmodern thinking not only elucidates the broader context of psychology as an academic discipline but rather that experimental psychology exhibits a number of aspects which can best be described as facets of postmodern thinking although they are not acknowledged as such.

The Postmodern Context of Experimental Psychology

Paradoxically, postmodern elements have been present since the very beginning of experimental psychology although postmodernism gained momentum only decades later. One of the characteristics of postmodernism is the transplantation of certain elements from their original context to new contexts, e.g., the popularity of "Eastern" philosophies and practices in contemporary "Western" societies. These different elements are often juxtaposed and combined to create something new, e.g., new "westernized" forms of yoga (Shearer, 2020).

Similarly, the founders of modern academic psychology took up the scientific method, which was originally developed in the context of the natural sciences, and transplanted it to the study of the human *psyche* in the hope to repeat the success of the natural sciences. By contrast, methods developed specifically in the context of psychology such as psychoanalysis (Wax, 1995) or introspection according to the Würzburg School (Hackert and Weger, 2018) have gained much less ground in academic psychology. The way we understand both the *psyche* and psychology has been shaped to a great extent by the transfer of the principles of modern science, namely quantitative measurement and experimental methods, although it is not evident *per se* that this is the best approach to elucidate mental and behavioral phenomena. Applying the methods of the natural sciences to a new and different context, namely to phenomena pertaining to the human *psyche*, is a truly postmodern endeavor because it juxtaposes two quite distinct areas and merges them into something new—experimental psychology.

The postmodern character of experimental psychology becomes evident on two levels: First, the subject matter—the human *psyche*—exhibits a postmodern character since mental and behavioral phenomena are highly dependent on the idiosyncratic contexts of the involved individuals, which makes it impossible to establish unambiguous general laws to describe them. Second, experimental psychology itself displays substantial postmodern traits because both its method and the knowledge it produces—although seemingly objective and rooted in the

modern scientific worldview—inevitably contain postmodern elements, as will be shown below.

The Experiment as Simulacrum

The term "simulacrum" basically means "copy," often in the sense of "inferior copy" or "phantasm/illusion." However, in postmodern usage "simulacrum" has acquired a more nuanced and concrete meaning. "Simulacrum" is a key term in the work of postmodern philosopher Jean Baudrillard, who arguably presented the most elaborate theory on simulacra (1981/1994). According to Baudrillard, a simulacrum "is the reflection of a profound ['real'] reality" (16/6). Simulacra, however, are more than identical carbon copies because they gain a life of their own and become "real" in the sense of becoming an own entity. For example, the personality a pop star shows on stage is not "real" in the sense that it is their "normal," off-stage personality, but it is certainly "real" in the sense that it is perceived by the audience even if they are aware that it might be an "artificial" personality. Two identical cars can also be "different" for one might be used as a means of transportation while the other might be a status symbol. Even an honest video documentation of a certain event is not simply a copy of the events that took place because it lies within the medium video that only certain sections can be recorded from a certain perspective. Additionally, the playback happens in other contexts as the original event, which may also alter the perception of the viewer.

The post-structuralist—an approach closely associated with postmodernism—philosopher Roland Barthes pointed out another important aspect of simulacra. He contended that in order to understand something—an "object" in Barthes' terminology—we necessarily create simulacra because we "*reconstruct* [our italics] an 'object' in such a way as to manifest thereby the rules of functioning [···] of this object" (Barthes, 1963, 213/214). In other words, when we investigate an object—any phenomenon, either material, mental, or social—we have to perceive it first. This means that we must have some kind of mental representation of the phenomenon/object—and it is crucial to note that this representation is not the same thing as the "real" object itself. All our mental operations are therefore not performed on the "real" object but on mental representations of the object. We decompose a phenomenon in order to understand it, that is, we try to identify its components. In doing so, we effect a change in the object because our phenomenon is no longer the original phenomenon "as it is" for we are performing a mental operation on it, thereby transforming the original phenomenon. Identifying components may be simple, e.g., dividing a tree into roots, trunk, branches, and leaves may seem obvious or even "natural" but it is nevertheless us as investigators who create this structure—the tree itself is probably not aware of it. Now that we have established this structure, we are able to say that the tree consists of several components and name these components. Thus, we have introduced "new" elements *into our understanding* of the tree. This is the important point, even though the elements, i.e., the branches and leaves themselves "as they are," have naturally always been "present." Our understanding of "tree" has therefore changed completely because a tree is now something which is composed of several elements. In that sense, we have changed the original phenomenon by adding something—and this has all happened in our thinking and not in the tree itself. It is also possible to find different structures and different components for the tree, e.g., the brown and the green, which shows that we construct this knowledge.

Next, we can investigate the components to see how they interact with and relate to each other and to the whole system. Also, we can work out their functions and determine the conditions under which a certain event will occur. We can even expand the scope of our investigation and examine the tree in the context of its ecosystem. But no matter what we do or how sophisticated our investigation becomes, everything said above remains true here, too, because neither all these actions listed above nor the knowledge we gain from them are the object itself. Rather, we have added something to the object and the more we know about our object, the more knowledge we have constructed. This addition is what science—gaining knowledge—is all about. Or in the words of Roland Barthes: "the simulacrum is intellect added to object, and this addition has an anthropological value, in that it is man himself, his history, his situation, his freedom and the very resistance which nature offers to his mind" (1963/1972, 214/215).

In principle, this holds truth regarding all scientific investigations. But the more complex phenomena are, the more effort and personal contribution is required on behalf of the investigator to come up with structures, theories, or explanations. Paraphrasing Barthes: When dealing with complex phenomena, more intellect must be added to the object, which means in turn that there are more possibilities for different approaches and perspectives, that is, the constructive element becomes larger. As discussed previously, this does not mean that investigative and interpretative processes are arbitrary. But it is clear from this train of thought that "objectivity" or "truth" in a "positivist," naïve empiricist "realist," or absolute sense are not attainable. Nevertheless, we argue here that this is not a drawback, as many critics of postmodernism contend (see above), but rather an advantage because it allows more accurate scientific investigations of true-to-life phenomena, which are typically complex in the case of psychology.

The concepts of simulacra by Baudrillard and Barthes can be combined to provide a description of the experiment in psychology. Accordingly, our understanding of the concept of the "simulacrum" entails that scientific processes—indeed all investigative processes—necessarily need to duplicate the object of their investigation in order to understand it. In doing so, constructive elements are necessarily introduced. These elements are of a varying nature, which means that investigations of one and the same phenomenon may differ from each other and different investigations may find out different things about the phenomenon in question. These investigations then become entities on their own—in the Baudrillardian sense—and therefore simulacra.

In a groundbreaking article on "the meaning and limits of exact science" physicist Max Planck stated that "[a]n experiment is a question which science poses to nature, and a measurement is the recording of nature's answer" (Planck, 1949, 325). The act of "asking a question" implies that the person asking the question

has at least a general idea of what the answer might look like (Heidegger, 1953, §2). For example: When asking someone for their name, we obviously do not know what they are called, but we assume that they have a name and we also have an idea of how the concept "name" works. Otherwise we could not even conceive, let alone formulate, and pose our question. This highlights how a certain degree of knowledge and understanding of a concept is necessary so that we are able to ask questions about it. Likewise, we need to have a principal idea or assumption of possible mechanisms if we want to find out how more complex phenomena function. It is—at least at the beginning—irrelevant whether these ideas are factually correct or entirely wrong, for without them we would be unable to approach our subject matter in the first place.

The context of the investigator—their general worldview, their previous knowledge and understanding, and their social situation—obviously plays an important part in the process of forming a question which can be asked in the current research context. Although this context may be analyzed along postmodern lines in order to find out how it affects research, production of knowledge, and—when the knowledge is applied—possible (social) consequences, there is a much more profound implication pertaining to the very nature of the experiment as a means to gain knowledge.

Irrespective of whether it is a simple experiment in physics such as Galileo Galilei's or an experiment on a complex phenomenon from social or cognitive psychology, the experiment is a situation which is specifically *designed* to answer a certain type of questions, usually causal relationships, such as: "Does A causally affect B?" Excluding the extremely complex discussion on the nature of causality and causation (e.g., Armstrong, 1997; Pearl, 2009; Paul and Hall, 2013), it is crucial to note that we need the experiment as a tool to answer this question. Although we may theorize about a phenomenon and infer causal relationships simply by observing, we cannot—at least according to the prevailing understanding of causality in the sciences—prove causal relationships without the experiment.

The basic idea of the experiment is to create conditions which differ in only one single factor which is suspected as a causal factor for an effect. The influence of all other potential causal relationships is kept identical because they are considered as confounding factors which are irrelevant from the perspective of the research question of the current experiment. Then, if a difference is found in the outcome between the experimental conditions, this is considered as proof that the aspect in question exerts indeed a causal effect. This procedure and the logic behind it are not difficult to understand. However, a closer look reveals that this is actually far from simple or obvious.

To begin with, an experiment is nothing which occurs "naturally" but a situation created for a specific purpose, i.e., an "artificial" situation, because other causal factors exerting influence in "real" life outside the laboratory are deliberately excluded and considered as "confounding" factors. This in itself shows that the experiment contains a substantial postmodern element because instead of creating something it rather *re*-creates it. This re-creation is of course based on phenomena from the "profound" reality—in the Baudrillardian sense—since the explicit aim is to find out something about this profound reality and not to create something new or something else. However, as stated above, this re-creation must contain constructive elements reflecting the presuppositions, conceptual-theoretical assumptions, and aims of the investigator. By focusing on one factor and by reducing the complexity of the profound reality, the practical operationalization and realization thus reflect both the underlying conceptual structure and the anticipated outcome as they are specifically designed to test for the suspected but hidden or obscured causal relationships.

At this point, another element becomes relevant, namely the all-important role of language, which is emphasized in postmodern thinking (e.g., Harris, 2005). Without going into the intricacies of semiotics, there is an explanatory gap (Chalmers, 2005)—to borrow a phrase from philosophy of mind—between the phenomenon on the one hand and the linguistic and/or mental representation of it on the other. This relationship is far from clear and it is therefore problematic to assume that our linguistic or mental representations—our words and the concepts they designate—are identical with the phenomena themselves. Although we cannot, at least according to our present knowledge and understanding, fully bridge this gap, it is essential to be aware of it in order to avoid some pitfalls, as will be shown in the examples below.

Even a seemingly simple word like "tree"—to take up once more our previous example—refers to a tangible phenomenon because there are trees "out there." However, they come in all shapes and sizes, there are different kinds of trees, and every single one of them may be labeled as "tree." Furthermore, trees are composed of different parts, and the leaf—although part of the tree—has its own word, i.e., linguistic and mental representation. Although the leaf is part of the tree—at least according to *our* concepts—it is unclear whether "tree" also somehow encompasses "leaf." The same holds true for the molecular, atomic, or even subatomic levels, where there "is" no tree. Excluding the extremely complex ontological implications of this problem, it has become clear that we are referring to a certain level of granularity when using the word "tree." The level of granularity reflects the context, aims, and concepts of the investigator, e.g., an investigation of the rain forest as an ecosystem will ignore the subatomic level.

How does this concern experimental psychology? Psychology studies intangible phenomena, namely mental and behavioral processes, such as cognition, memory, learning, motivation, emotion, perception, consciousness, etc. It is important to note that these terms designate *theoretical* constructs as, for example, memory cannot be observed directly. We may provide the subjects of an experiment a set of words to learn and observe later how many words they reproduce correctly. A theoretical construct therefore describes such relationships between stimulus and behavior, and we may *draw conclusions* from this observable data *about* memory. But neither the observable behavior of the subject, the resulting data, nor our conclusions are identical with memory itself.

This train of thought demonstrates the postmodern character of experimental psychology because we construct our knowledge. But there is more to it than that: Even by trying to define

a theoretical construct as exactly as possible—e.g., memory as "the process of maintaining information over time" (Matlin, 2012, 505) or "the means by which we retain and draw on our past experiences to use this information in the present" (Sternberg and Sternberg, 2011, 187)—the explanatory gap between representation and phenomenon cannot be bridged. Rather, it becomes even more complicated because theoretical constructs are composed of other theoretical constructs, which results in some kind of self-referential circularity where constructs are defined by other constructs which refer to further constructs. In the definitions above, for instance, hardly any key term is self-evident and unambiguous for there are different interpretations of the constructs "process," "maintaining," "information," "means," "retain," "draw on," "experiences," and "use" according to their respective contexts. Only the temporal expressions "over time," "past," and "present" are probably less ambiguous here because they are employed as non-technical, everyday terms. However, the definitions above are certainly not entirely incomprehensible—in fact, they are rather easy to understand in everyday language—and it is *quite clear* what the *authors intend* to *express*. The italics indicate constructive elements, which demonstrates that attempts to give a precise definition in the language of science result in fuzziness and self-reference.

Based on a story by Jorge Luis Borges, Baudrillard (1981) found an illustrative allegory: a map so precise that it portrays everything in perfect detail—but therefore inevitably so large that it shrouds the entire territory it depicts. Similarly, Taleb (2007) coined the term "ludic fallacy" for mistaking the model/map—in our context: experiments in psychology—for the reality/territory, that is, a mental or behavioral phenomenon. Similar to the functionality of a seemingly "imprecise" map which contains only the relevant landmarks so the user may find their way, the fuzziness of language poses no problems in everyday communication. So why is it a problem in experimental psychology? Since the nature of theoretical constructs in psychology lies precisely in their very fuzziness, the aim of reaching a high degree of granularity and precision in experimental psychology seems to be unattainable (see the various failed attempts to create "perfect" languages which might depict literally everything "perfectly," e.g., Carapezza and D'Agostino, 2010).

Without speculating about ontic or epistemic implications, it is necessary to be aware of the explanatory gap and to refrain from identifying the experiment and the underlying operationalization with the theoretical construct. Otherwise, this gap is "filled" unintentionally and uncontrollably if the results of an experiment are taken as valid proof for a certain theoretical construct, which is actually fuzzy and potentially operationalizable in a variety of ways. If this is not acknowledged, words, such as "memory," become merely symbols devoid of concrete meaning, much like a glass bead game—or in postmodern terminology: a hyperreality.

Experiments and Hyperreality

"Hyperreality" is another key term in the work of Jean Baudrillard (1981) and it denotes a concept closely related to the simulacrum. Accordingly, in modern society the simulacra are ubiquitous and they form a system of interconnected simulacra which refer to each other rather than to the real, thereby possibly hiding or replacing the real. Consequently, the simulacra become real in their own right and form a "more real" reality, namely the hyperreality. One may or may not accept Baudrillard's conception, especially the all-embracing social and societal implications, but the core concept of "hyperreality" is nevertheless a fruitful tool to analyze experimental psychology. We have already seen that the experiment displays many characteristics of a simulacrum, so it is not surprising that the concept of hyperreality is applicable here as well, although in a slightly different interpretation than Baudrillard's.

The hyperreal character of the experiment can be discussed on two levels: the experiment itself and the discourse wherein it is embedded.

On the level of the experiment itself, two curious observations must be taken into account. First, and in contrast to the natural sciences where the investigator is human and the subject matter (mostly) non-human and usually inanimate, in psychology both the investigator and the subject matter are human. This means that the subjects of the experiment, being autonomous persons, are not malleable or completely controllable by the investigator because they bring their own background, history, worldview, expectations, and motivations. They interpret the situation—the experiment—and act accordingly, but not necessarily in the way the investigator had planned or anticipated (Smedslund, 2016). Therefore, the subjects create their own versions of the experiment, or, in postmodern terminology, a variety of simulacra, which may be more or less compatible with the framework of the investigator. This holds true for all subjects of an experiment, which means that the experiment as a whole may also be interpreted as an aggregation of interconnected simulacra—a hyperreality.

The hyperreal character becomes even more evident because what contributes in the end to the interpretation of the results of the experiment are not the actual performances and results of the individual subjects as they were intended by them but rather how their performances and results are handled, seen, and interpreted by the investigator. Even if the investigator tries to be as faithful as possible and aims at an exact and unbiased measurement—i.e., an exact copy—there are inevitably constructive elements which introduce uncertainty into the experiment. Investigators can never be certain what the subjects were actually doing and thinking so they must necessarily work with interpretations. Or in postmodern terms: Because the actual performances and results of the subjects are not directly available the investigators must deal with simulacra. These simulacra become the investigators' reality and thus any further treatment—statistical analyses, interpretations, or discussions—becomes a hyperreality, that is, a set of interconnected simulacra which have become "real."

On the level of the discourse wherein the experiment is embedded, another curious aspect also demonstrates the hyperreal character of experimental psychology. Psychology is, according to the standard definition, the scientific study of mental and behavioral processes of the individual (e.g., Gerrig, 2012). This definition contains two actually contradictory elements. On the one hand, the focus is on processes of the individual.

On the other hand, the—scientific—method to elucidate these processes does not look at individuals *per se* but aggregates their individual experiences and transforms them into a "standard" experience. The results from experiments, our knowledge of the human psyche, reflect psychological functioning at the level of the mean across individuals. And even if we assume that the mean is only an estimator and not an exact description or prediction, the question remains open how de-individualized observations are related to the experience of an individual. A general mechanism, a law—which was discovered by abstracting from a multitude of individual experiences—is then (*re*-)imposed in the opposite direction back onto the individual. In other words, a simulacrum—namely, the result of an experiment—is viewed and treated as reality, thus becoming hyperreal. Additionally, and simply because it is considered universally true, this postulated law acquires thereby a certain validity and "truth"—often irrespective of its actual, factual, or "profound" truth—on its own. Therefore, it can become impossible to distinguish between "profound" and "simulacral" truth, which is the hallmark of hyperreality.

Measuring the Capacity of the Visual Working Memory

Vision is an important sensory modality and there is extensive research on this area (Hutmacher, 2019). Much of our daily experience is shaped by seeing a rich and complex world around us, and it is therefore an interesting question how much visual information we can store and process. Based on the development of a seminal experimental paradigm, Luck and Vogel (1997) have shown that visual working memory has a storage capacity of about four items. This finding is reported in many textbooks (e.g., Baddeley, 2007; Parkin, 2013; Goldstein, 2015) and has almost become a truism in cognitive psychology.

The experimental paradigm developed by Luck and Vogel (1997) is a prime example of an experiment which closely adheres to the scientific principles outlined above. In order to make a very broad and fuzzy phenomenon measurable, simple abstract forms are employed as visual stimuli—such as colored squares, triangles, or lines, usually on a "neutral," e.g., gray, background—which can be counted in order to measure the capacity of visual working memory. Reducing the exuberant diversity of the "outside visual world" to a few abstract geometric forms is an extremely artificial situation. The obvious contrast between simple geometrical forms and the rich panorama of the "real" visual world illustrates the pitfalls of controlling supposed confounding variables, namely the incontrollable variety of the "real" world and how we see it. Precisely by abstracting and by excluding potential confounding variables it is possible to count the items and to make the capacity of the visual working memory measurable. But in doing so the original phenomenon—seeing the whole world—is lost. In other words: A simulacrum has been created.

The establishment of the experimental paradigm by Luck and Vogel has led to much research and sparked an extensive discussion how the limitation to only four items might be explained (see the summaries by Brady et al., 2011; Luck and Vogel, 2013; Ma et al., 2014; Schurgin, 2018). However, critically, several studies have shown that the situation is different when real-world objects are used as visual stimuli rather than simple abstract forms, revealing that the capacity of the visual working memory is higher for real-world objects (Endress and Potter, 2014; Brady et al., 2016; Schurgin et al., 2018; Robinson et al., 2020; also Schurgin and Brady, 2019). Such findings show that the discourse about the mechanisms behind the limitations of the visual working memory is mostly about an artificial phenomenon which has no counterpart in "reality"—the perfect example of a hyperreality.

This hyperreal character does not mean that the findings of Luck and Vogel (1997) or similar experiments employing artificial stimuli are irrelevant or not "true." The results are true—but it is a *local* truth, only valid for the specific context of *specific* experiments, and not a *global* truth which applies to the visual working memory *in general*. That is, speaking about "visual working memory" based on the paradigm of Luck and Vogel is a mistake because it is actually about "visual working memory for simple abstract geometrical forms in front of a gray background."

Free Will and Experimental Psychology

The term "free will" expresses the idea of having "a significant kind of *control* [italics in the original] over one's actions" (O'Connor and Franklin, 2018, n.p.). This concept has occupied a central position in Western philosophy since antiquity because it has far-reaching consequences for our self-conception as humans and our position in the world, including questions of morality, responsibility, and the nature of legal systems (e.g., Beebee, 2013; McKenna and Pereboom, 2016; O'Connor and Franklin, 2018). Being a topic of general interest, it is not surprising that experimental psychologists have tried to investigate free will as well.

The most famous study was conducted by Libet et al. (1983), and this experiment has quickly become a focal point in the extensive discourse on free will because it provides empirical data and a scientific investigation. Libet et al.'s experiment seems to show that the subjective impression when persons consciously decide to act is in fact preceded by objectively measurable but unconscious physical processes. This purportedly proves that our seemingly voluntary actions are actually predetermined by physical processes because the brain has unconsciously reached a decision already before the person becomes aware of it and that our conscious intentions are simply grafted onto it. Therefore, we do not have a free will, and consequently much of our social fabric is based on an illusion. Or so the story goes.

This description, although phrased somewhat pointedly, represents a typical line of thought in the discourse on free will (e.g., the prominent psychologists Gazzaniga, 2011; Wegner, 2017; see Kihlstrom, 2017, for further examples).

Libet's experiment sparked an extensive and highly controversial discussion: For some authors, it is a refutation or at least threat to various concepts of free will, or, conversely, an indicator or even proof for some kind of material determinism. By contrast, other authors deny that the experiment refutes or counts against free will. Furthermore, a third group—whose position we adopt for our further argumentation—denies that

Libet's findings are even relevant for this question at all (for summaries of this complex and extensive discussion and various positions including further references see Nahmias, 2010; Radder and Meynen, 2013; Schlosser, 2014; Fischborn, 2016; Lavazza, 2016; Schurger, 2017). Libet's own position, although not entirely consistent, opposes most notions of free will (Roskies, 2011; Seifert, 2011). Given this background, it is not surprising that there are also numerous further experimental studies on various aspects of this subject area (see the summaries by Saigle et al., 2018; Shepard, 2018; Brass et al., 2019).

However, we argue that this entire discourse is best understood along postmodern lines as hyperreality and that Libet's experiment itself is a perfect example of a simulacrum. A closer look at the concrete procedure of the experiment shows that Libet actually asked his participants to move their hand or finger "at will" while their brain activity was monitored with an EEG. They were instructed to keep watch in an introspective manner for the moment when they felt the "urge" to move their hand and to record this moment by indicating the clock-position of a pointer. This is obviously a highly artificial situation where the broad and fuzzy concept of "free will" is abstracted and reduced to the movement of the finger, the only degree of freedom being the moment of the movement. The question whether this is an adequate operationalization of free will is of paramount importance, and there are many objections that Libet's setup fails to measure free will at all (e.g., Mele, 2007; Roskies, 2011; Kihlstrom, 2017; Brass et al., 2019).

Before Libet, there was no indication that the decision when to move a finger might be relevant for the concept of free will and the associated discourse. The question whether we have control over our actions referred to completely different levels of granularity. Free will was discussed with respect to questions such as whether we are free to live our lives according to our wishes or whether we are responsible for our actions in social contexts (e.g., Beebee, 2013; McKenna and Pereboom, 2016; O'Connor and Franklin, 2018), and not whether we lift a finger now or two seconds later. Libet's and others' jumping from very specific situations to far-reaching conclusions about a very broad and fuzzy theoretical construct illustrates that an extremely wide chasm between two phenomena, namely moving the finger and free will, is bridged in one fell swoop.

In other words, Libet's experiment is a simulacrum as it duplicates a phenomenon from our day-to-day experience—namely free will—but in doing so the operationalization alters and reduces the theoretical construct. The outcome is a questionable procedure whose relationship to the phenomenon is highly controversial. Furthermore, the fact that, despite its tenuous connection to free will, Libet's experiment sparked an extensive discussion on this subject reveals the hyperreal nature of the entire discourse because what is being discussed is not the actual question—namely free will—but rather a simulacrum. Everything else—the arguments, counter-arguments, follow-up experiments, and their interpretations—built upon Libet's experiment are basically commentaries to a simulacrum and not on the real phenomena. Therefore, a hyperreality is created where the discourse revolves around entirely artificial phenomena, but where the arguments in this discussion refer back to and affect the real as suggestions are made to alter the legal system and our ideas of responsibility—which, incidentally, is not a question of empirical science but of law, ethics, and philosophy.

All of the above is not meant to say that this whole discourse is meaningless or even gratuitous—on the contrary, our understanding of the subject matter has greatly increased. Although our knowledge of free will has hardly increased, we have gained much insight into the hermeneutics and methodology—and pitfalls!—of *investigations* of free will, possible consequences on the individual and societal level, and the workings of scientific discourses. And this is exactly what postmodernism is about.

DISCUSSION

As shown above, there are a number of postmodern elements in the practice of experimental psychology: The prominent role of language, the gap between the linguistic or mental representation and the phenomenon, the "addition of intellect to the object," the simulacral character of the experiment itself in its attempt to re-create phenomena, which necessarily transforms the "real" phenomenon due to the requirements of the experiment, and finally the creation of a hyperreality if experiments are taken as the "real" phenomenon and the scientific discourse becomes an exchange of symbolic expressions referring to the simulacra created in experiments, replacing the real. All these aspects did not seep gradually into experimental psychology in the wake of postmodernism but have been present since the very inception of experimental psychology as they are necessarily inherent in its philosophy of science.

Given these inherent postmodern traits in experimental psychology, it is puzzling that there is so much resistance against a perceived "threat" of psychology's scientificness. Although a detailed investigation of the reasons lies outside the scope of this analysis, we suspect there are two main causes: First, an insufficient knowledge of the history of science and understanding of philosophy of science may result in idealized concepts of a "pure" natural science. Second, lacking familiarity with basic tenets of postmodern approaches may lead to the assumption that postmodernism is just an idle game of arbitrary words. However, "science" and "postmodernism" and their respective epistemological concepts are not opposites (Gergen, 2001; Holtz, 2020). This is especially true for psychology, which necessarily contains a social dimension because not only the investigators are humans but also the very subject matter itself.

The (over-)reliance on quantitative-experimental methods in psychology, often paired with a superficial understanding of the philosophy of science behind it, has been criticized, either from the theoretical point of view (e.g., Bergmann and Spence, 1941; Hearnshaw, 1941; Petrie, 1971; Law, 2004; Smedslund, 2016) or because the experimental approach has failed to produce reliable, valid, and relevant applicable knowledge in educational psychology (Slavin, 2002). It is perhaps symptomatic that a textbook teaching the principles of science for psychologists does not contain even one example from experimental psychology but employs only examples from physics, plus Darwin's theory of evolution (Wilton and Harley, 2017).

On the other hand, the postmodern perspective on experimental psychology provides insight into some pitfalls, as illustrated by the examples above. On the level of the experiment, the methodological requirements imply the creation of an artificial situation, which opens up a gap between the phenomenon as it is in reality and as it is concretely operationalized in the experimental situation. This is not a problem *per se* as long as is it clear—and clearly communicated!—that the results of the experiment are only valid in a certain context. The problems begin if the movement of a finger is mistaken for free will. Similarly, being aware that local causalities do not explain complex phenomena such as mental and behavioral processes in their entirety also prevents (over-) generalization, especially if communicated appropriately. These limitations make it clear that the experiment should not be made into an absolute or seen as the only valid way of understanding the *psyche* and the world.

On the level of psychology as an academic discipline, any investigation must select the appropriate level of granularity and strike a balance between the methodological requirements and the general meaning of the theoretical concept in question to find out something about the "real" world. If the level of granularity is so fine that results cannot be tied back to broader theoretical constructs rather than providing a helpful understanding of our psychological functioning, academic psychology is in danger of becoming a self-referential hyperreality.

The postmodern character of experimental psychology also allows for a different view on the so-called replication crisis in psychology. Authors contending that there is no replication crisis often employ arguments which exhibit postmodern elements, such as the emphasis on specific local conditions in experiments which may explain different outcomes of replication studies (Stroebe and Strack, 2014; Baumeister, 2019). In other words, they invoke the simulacral character of experiments. This explanation may be valid or not, but the replication crisis has shown the limits of a predominantly experimental approach in psychology.

Acknowledging the postmodern nature of experimental psychology and incorporating postmodern thinking explicitly into our research may offer a way out of this situation. Our subject matter—the *psyche*—is extremely complex, ambiguous, and often contradictory. And postmodern thinking has proven capable of successfully explaining such phenomena (e.g., Bertens, 1995; Sim, 2011; Aylesworth, 2015). Thus, paradoxically, by accepting and considering the inherently fuzzy nature of theoretical constructs, they often become much clearer (Ronzitti, 2011). Therefore, thinking more along postmodern lines in psychology would actually sharpen the theoretical and conceptual basis of experimental psychology—all the more as experimental psychology has inevitably been a postmodern endeavor since its very beginning.

AUTHOR CONTRIBUTIONS

RM, CK, and CL developed the idea for this article. RM drafted the manuscript. CK and CL provided feedback and suggestions. All authors approved the manuscript for submission.

REFERENCES

Armstrong, D. M. (1997). *A World of States of Affairs*. Cambridge: CUP.

Aylesworth, G. (2015). "Postmodernism," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta. Available online at: https://plato.stanford.edu/entries/postmodernism/

Baddeley, A. (2007). *Working Memory, Thought, and Action*. Oxford: OUP.

Barthes, R. (1963). "L'activité structuraliste," in *Essais Critiques* (pp. 215–218). Paris: Éditions du Seuil. ["Structuralist activity." Translated by R. Howard (1972). In *Critical Essays*, ed. R. Barthes (Evanston: Northern University Press), 213–220].

Baudrillard, J. (1981). *Simulacres et Simulation*. Paris: Galilée. [*Simulacra and Simulation*. Translated by S. F. Glaser (1994). Ann Arbor: The University of Michigan Press.]

Baumeister, R. F. (2019). "Self-control, ego depletion, and social psychology's replication crisis," in *Surrounding Self-control* (Appendix to chap. 2), ed. A. Mele (New York, NY: OUP). doi: 10.31234/osf.io/uf3cn

Beebee, H. (2013). *Free Will: An Introduction*. New York, NY: Palgrave Macmillan.

Bergmann, G., and Spence, K. W. (1941). Operationism and theory in psychology. *Psychol. Rev.* 48, 1–14. doi: 10.1037/h0054874

Bertens, H. (1995). *The Idea of the Postmodern. A History*. London: Routledge.

Brady, T. F., Konkle, T., and Alvarez, G. A. (2011). A review of visual memory capacity: beyond individual items and toward structured representations. *J. Vis.* 11, 1–34. doi: 10.1167/11.5.4

Brady, T. F., Störmer, V. S., and Alvarez, G. A. (2016). Working memory is not fixed-capacity: more active storage capacity for real-world objects than for simple stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 113, 7459–7464. doi: 10.1073/pnas.1520027113

Brass, M., Furstenberg, A., and Mele, A. R. (2019). Why neuroscience does not disprove free will. *Neurosci. Biobehav. Rev.* 102, 251–263. doi: 10.1016/j.neubiorev.2019.04.024

Capaldi, E. J., and Proctor, R. W. (2013). "Postmodernism and the development of the psychology of science," in *Handbook of the Psychology of Science*, eds G. J. Feist and M. E. Gorman (New York, NY: Springer), 331–352.

Carapezza, M., and D'Agostino, M. (2010). Logic and the myth of the perfect language. *Logic Philos. Sci.* 8, 1–29. doi: 10.1093/oso/9780190869816.003.0001

Chalmers, D. (2005). "Phenomenal concepts and the explanatory gap," in *Phenomenal Concepts and Phenomenal Knowledge. New Essays on Consciousness and Physicalism*, eds T. Alter and S. Walter (Oxford: OUP), 167–194. doi: 10.1093/acprof:oso/9780195171655.003.0009

Cohen, H. F. (2015). *The Rise of Modern Science Explained: A Comparative History*. Cambridge: CUP.

Endress, A. D., and Potter, M. C. (2014). Large capacity temporary visual memory. *J. Exp. Psychol. Gen.* 143, 548–565. doi: 10.1037/a0033934

Fischborn, M. (2016). Libet-style experiments, neuroscience, and libertarian free will. *Philos. Psychol.* 29, 494–502. doi: 10.1080/09515089.2016.1141399

Gazzaniga, M. S. (2011). *Who's in Charge? Free Will and the Science of the Brain*. New York, NY: Ecco.

Gemignani, M., and Peña, E. (2007). Postmodern conceptualizations of culture in social constructionism and cultural studies. *J. Theor. Philos. Psychol.* 27–28, 276–300. doi: 10.1037/h0091297

Gergen, K. J. (2001). Psychological science in a postmodern context. *Am. Psychol.* 56, 803–813. doi: 10.1037/0003-066X.56.10.803

Gergen, K. J. (2002). Psychological science: to conserve or create? *Am. Psychol.* 57, 463–464. doi: 10.1037/0003-066X.57.6-7.463

Gerrig, R. J. (2012). *Psychology and Life*, 20th Edn. Boston: Pearson.

Goldstein, E. B. (2015). *Cognitive Psychology: Connecting Mind, Research and Everyday Experience*. Stamford: Cengage Learning.

Goodwin, C. J. (2015). *A History of Modern Psychology*, 5th Edn. Hoboken, NJ: Wiley.

Hackert, B., and Weger, U. (2018). Introspection and the Würzburg school: implications for experimental psychology today. *Eur. Psychol.* 23, 217–232. doi: 10.1027/1016-9040/a000329

Haig, B. D. (2014). *Investigating the Psychological World: Scientific Method in the Behavioral Sciences*. Cambridge, MA: MIT Press.

Hansen, J. T. (2015). The relevance of postmodernism to counselors and counseling practice. *J. Ment. Health Counsel.* 37, 355–363. doi: 10.17744/MEHC.37.4.06

Hare-Mustin, R. T., and Marecek, J. (1988). The meaning of difference: gender theory, postmodernism, and psychology. *Am. Psychol.* 43, 455–464. doi: 10.1037//0003-066X.43.6.455

Harris, R. (2005). *The Semantics of Science*. London: Continuum.

Hearnshaw, L. S. (1941). Psychology and operationism. *Aust. J. Psychol. Philos.* 19, 44–57. doi: 10.1080/00048404108541506

Heidegger, M. (1953). *Sein und Zeit* (7. Aufl.). Tübingen: Niemeyer. [*Being and Time*. Translated by J. Stambaugh, revised by D. J. Schmidt (2010). Albany: SUNY Press.]

Held, B. S. (2007). *Psychology's Interpretive Turn: The Search for Truth and Agency in Theoretical and Philosophical Psychology*. Washington, DC: APA.

Henry, J. (1997). *The Scientific Revolution and the Origins of Modern Science*. Basingstoke: Macmillan.

Hodge, B. (1999). The Sokal 'Hoax': some implications for science and postmodernism. *Continuum J. Media Cult. Stud.* 13, 255–269. doi: 10.1080/10304319909365797

Hofmann, S. G. (2002). More science, not less. *Am. Psychol.* 57:462. doi: 10.1037//0003-066X.57.6-7.462a

Holt, R. R. (2005). "The menace of postmodernism to a psychoanalytic psychology," in *Relatedness, Self-definition and Mental Representation: Essays in Honor of Sidney J. Blatt*, eds J. S. Auerbach, K. N. Levy, and C. E. Schaffer (London: Routledge), 288–302. doi: 10.4324/9780203337318_chapter_18

Holtz, P. (2020). Does postmodernism really entail a disregard for the truth? Similarities and differences in postmodern and critical rationalist conceptualizations of truth, progress, and empirical research methods. *Front. Psychol.* 11:545959. doi: 10.3389/fpsyg.2020.545959

Holzman, L. (2006). Activating postmodernism. *Theory Psychol.* 16, 109–123. doi: 10.1177/0959354306060110

Holzman, L., and Morss, J. (eds). (2000). *Postmodern Psychologies, Societal Practice, and Political Life*. New York, NY: Routledge.

Hutmacher, F. (2019). Why is there so much more research on vision than on any other sensory modality? *Front. Psychol.* 10:2246. doi: 10.3389/fpsyg.2019.02246

Jager, B. (1991). Psychology in a postmodern era. *J. Phenomenol. Psychol.* 22, 60–71. doi: 10.1163/156916291X00046

Jiménez, J. P. (2015). Psychoanalysis in postmodern times: some questions and challenges. *Psychoanal. Inquiry* 35, 609–624. doi: 10.1080/07351690.2015.1055221

Kihlstrom, J. F. (2017). Time to lay the Libet experiment to rest: commentary on Papanicolaou (2017). *Psycho. Conscious. Theory Res. Pract.* 4, 324–329. doi: 10.1037/cns0000124

Kleinert, A. (2009). Der messende Luchs. *NTM. Z. Gesch. Wiss. Tech. Med.* 17, 199–206. doi: 10.1007/s00048-009-0335-4

Krippner, S. (2001). "Research methodology in humanistic psychology in the light of postmodernity," in *The Handbook of Humanistic Psychology: Leading Edges in Theory, Research, and Practice*, eds K. J. Schneider, J. F. Bugental, and J. F. Pierson (Thousand Oaks, CA: SAGE Publications), 290–304. doi: 10.4135/9781412976268.n22

Kruger, D. J. (2002). The deconstruction of constructivism. *Am. Psychol.* 57, 456–457. doi: 10.1037/0003-066X.57.6-7.456

Kvale, S. (ed.) (1992). *Psychology and Postmodernism*. London: SAGE.

Lavazza, A. (2016). Free will and neuroscience: from explaining freedom away to new ways of operationalizing and measuring it. *Front. Hum. Neurosci.* 10:262. doi: 10.3389/fnhum.2016.00262

Law, J. (2004). *After Method: Mess in Social Science Research*. London: Routledge.

Leahey, T. H. (2017). *A History of Psychology: From Antiquity to Modernity*, 8th Edn. New York, NY: Routledge.

Leffert, M. (2007). A contemporary integration of modern and postmodern trends in psychoanalysis. *J. Am. Psychoanal. Assoc.* 55, 177–197. doi: 10.1177/00030651070550011001

Libet, B., Gleason, C. A., Wright, E. W., and Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain* 106, 623–642. doi: 10.1093/brain/106.3.623

Lilienfeld, S. O. (2010). Can psychology become a science? *Pers. Individ. Differ.* 49, 281–288. doi: 10.1016/j.paid.2010.01.024

Locke, E. A. (2002). The dead end of postmodernism. *Am. Psychol.* 57:458. doi: 10.1037/0003-066X.57.6-7.458a

Luck, S. J., and Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281. doi: 10.1038/36846

Luck, S. J., and Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends Cogn. Sci.* 17, 391–400. doi: 10.1016/j.tics.2013.06.006

Ma, W. J., Husain, M., and Bays, P. M. (2014). Changing concepts of working memory. *Nature Neurosci.* 17, 347–356. doi: 10.1038/nn.3655

Mandler, G. (2007). *A History of Modern Experimental Psychology: From James and Wundt to Cognitive Science*. Cambridge, MA: MIT Press.

Martin, J. (2003). Positivism, quantification and the phenomena of psychology. *Theory Psychol.* 13, 33–38. doi: 10.1177/0959354303013001760

Martin, J., and Sugarman, J. (2009). Middle-ground theorizing, realism, and objectivity in psychology: a commentary on Held (2007). *Theory Psychol.* 19, 115–122. doi: 10.1177/0959354308101422

Matlin, M. W. (2012). *Cognition*, 8th Edn. Hoboken: Wiley.

McKenna, M., and Pereboom, D. (2016). *Free Will: A Contemporary Introduction*. New York, NY: Routledge.

Mele, A. R. (2007). "Decisions, intentions, urges, and free will: why libet has not shown what he says he has," in *Causation and Explanation*, eds J. K. Campbell, M. O'Rourke, and H. S. Silverstein (Cambridge, MA: MIT Press), 241–263.

Michell, J. (2003). The quantitative imperative: positivism, naïve realism and the place of qualitative methods in psychology. *Theory Psychol.* 13, 5–31. doi: 10.1177/0959354303013001758

Nahmias, E. (2010). "Scientific challenges to free will," in *A Companion to the Philosophy of Action*, eds C. Sandis and T. O'Connor (Malden: Wiley-Blackwell), 345–310. doi: 10.1002/9781444323528.ch44

Nola, R., and Sankey, H. (2014). *Theories of Scientific Method. An Introduction*. Stocksfield: Acumen.

O'Connor, T., and Franklin, C. (2018). "Free will," in *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta. Available online at: https://plato.stanford.edu/entries/freewill/

Olby, R. C., Cantor, G. N., Christie, J. R. R., and Hodge, M. J. S. (eds). (1991). *Companion to the History of Modern Science*. London: Routledge.

Osterlind, S. J. (2019). *The Error of Truth: How History and Mathematics Came Together to Form Our Character and Shape Our Worldview*. Oxford: OUP.

Parkin, A. J. (2013). *Essential Cognitive Psychology* (classic edition). London: Psychology Press.

Paul, L. A., and Hall, N. (2013). *Causation: A User's Guide*. Oxford: OUP.

Pearl, J. (2009). *Causality. Models, Reasoning, and Inference*, 2nd Edn. Cambridge: CUP.

Petrie, H. G. (1971). A dogma of operationalism in the social sciences. *Philos. Soc. Sci.* 1, 145–160. doi: 10.1177/004839317100100109

Pinker, S. (2018). *Enlightenment Now. The Case for Reason, Science, Humanism, and Progress*. New York, NY: Viking.

Planck, M. (1949). The meaning and limits of exact science. *Science* 110, 319–327. doi: 10.1126/science.110.2857.319

Radder, H., and Meynen, G. (2013). Does the brain "initiate" freely willed processes? A philosophy of science critique of Libet-type experiments and their interpretation. *Theory Psychol.* 23, 3–21. doi: 10.1177/0959354312460926

Ramey, H. L., and Grubb, S. (2009). Modernism, postmodernism and (evidence-based) practice. *Contemp. Fam. Ther.* 31, 75–86. doi: 10.1007/s10591-009-9086-6

Robinson, M. M., Benjamin, A. S., and Irwin, D. E. (2020). Is there a K in capacity? Assessing the structure of visual short-term memory. *Cogn. Psychol.* 121, 101305. doi: 10.1016/j.cogpsych.2020.101305

Ronzitti, G. (ed.) (2011). *Vagueness: A Guide*. Dordrecht: Springer.

Roskies, A. L. (2011). "Why Libet's studies don't pose a threat to free will," in *Conscious Will and Responsibility*, eds W. Sinnott-Armstrong and L. Nadel (Oxford: OUP), 11–22. doi: 10.1093/acprof:oso/9780195381641.003.0003

Saigle, V., Dubljević, V., and Racine, E. (2018). The impact of a landmark neuroscience study on free will: a qualitative analysis of articles using Libet and colleagues' methods. *AJOB Neurosci.* 9, 29–41. doi: 10.1080/21507740.2018.1425756

Schlosser, M. E. (2014). The neuroscientific study of free will: a diagnosis of the controversy. *Synthese* 191, 245–262. doi: 10.1007/s11229-013-0312-2

Schurger, A. (2017). "The neuropsychology of conscious volition," in *The Blackwell Companion to Consciousness*, eds S. Schneider and M. Velmans (Malden: Wiley Blackwell), 695–710. doi: 10.1002/9781119132363.ch49

Schurgin, M. W. (2018). Visual memory, the long and the short of it: a review of visual working memory and long-term memory. *Attention Percept. Psychophys.* 80, 1035–1056. doi: 10.3758/s13414-018-1522-y

Schurgin, M. W., and Brady, T. F. (2019). When "capacity" changes with set size: ensemble representations support the detection of across-category changes in visual working memory. *J. Vis.* 19, 1–13. doi: 10.1167/19.5.3

Schurgin, M. W., Cunningham, C. A., Egeth, H. E., and Brady, T. F. (2018). Visual long-term memory can replace active maintenance in visual working memory. *bioRxiv* [Preprint]. doi: 10.1101/381848

Segerstråle, U.C.O. (ed.) (2000). *Beyond the Science Wars: The Missing Discourse about Science and Society*. Albany: SUNY Press.

Seifert, J. (2011). In defense of free will: a critique of Benjamin Libet. *Rev. Metaphys.* 65, 377–407.

Shearer, A. (2020). *The Story of Yoga: From Ancient India to the Modern West*. London: Hurst & Company.

Shepard, J. (2018). How libet-style experiments may (or may not) challenge lay theories of free will. *AJOB Neurosci.* 9, 45–47. doi: 10.1080/21507740.2018.1425766

Sim, S. (2011). *The Routledge Companion to Postmodernism*, 3rd Edn. London: Routledge.

Sinacore, A. L., and Enns, C. Z. (2005). "Diversity feminisms: postmodern, women-of-color, antiracist, lesbian, third-wave, and global perspectives," in *Teaching and Social Justice: Integrating Multicultural and Feminist Theories in the Classroom*, eds C. Z. Enns and A. L. Sinacore (Washington, DC: APA), 41–68. doi: 10.1037/10929-003

Slavin, R. E. (2002). Evidence-based education policies: transforming educational practice and research. *Educ. Res.* 31, 15–21. doi: 10.3102/0013189X031007015

Smedslund, J. (2016). Why psychology cannot be an empirical science. *Integrative Psychol. Behav. Sci.* 50, 185–195. doi: 10.1007/s12124-015-9339-x

Sokal, A. D. (1996). A physicist experiments with cultural studies. *Lingua Franca* 6, 62–64.

Sternberg, R. J., and Sternberg, K. (2011). *Cognitive Psychology*, 6th Edn. Wadsworth: Cengage Learning.

Stroebe, W., and Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspect. Psychol. Sci.* 9, 59–71. doi: 10.1177/1745691613514450

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable*. New York, NY: Random House.

Tarescavage, A. M. (2020). Science Wars II: the insidious influence of postmodern ideology on clinical psychology (commentary on "Implications of ideological bias in social psychology on clinical practice"). *Clin. Psychol. Sci. Pract.* 27:e12319. doi: 10.1111/cpsp.12319

Teo, T. (2018). *Outline of Theoretical Psychology*. London: Palgrave Macmillan.

Wax, M. L. (1995). Method as madness science, hermeneutics, and art in psychoanalysis. *J. Am. Acad. Psychoanal.* 23, 525–543. doi: 10.1521/jaap.1.1995.23.4.525

Wegner, D. M. (2017). *The Illusion of Conscious Will*, 2nd Edn. Cambridge, MA: MIT Press.

Wilton, R., and Harley, T. (2017). *Science and Psychology*. London: Routledge.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 13 January 2021
doi: 10.3389/fpsyg.2020.603617

Quantification in Experimental Psychology and Pragmatic Epistemology: Tension Between the Scientific Imperative and the Social Imperative

*Hervé Guyon**

Université de Bretagne Occidentale, AMURE Laboratory, Brest, France

Keywords: bio-power, epistemology, measurement, ontology, pragmatism, quantitative psychology, statistical positivism

OPEN ACCESS

Edited by:
Peter Adriaan Edelsbrunner, ETH Zürich, Switzerland

Reviewed by:
Michael Smithson, Australian National University, Australia
Luca Mari, University Carlo Cattaneo, Italy

***Correspondence:**
Hervé Guyon
herve.guyon@univ-brest.fr

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *07 September 2020*
Accepted: *07 December 2020*
Published: *13 January 2021*

Citation:
Guyon H (2021) Quantification in Experimental Psychology and Pragmatic Epistemology: Tension Between the Scientific Imperative and the Social Imperative. Front. Psychol. 11:603617. doi: 10.3389/fpsyg.2020.603617

INTRODUCTION

This article is an opinion article that criticizes the usual practice in quantitative psychology. Our development seeks to link a (pragmatic) critique of measurement and statistical modeling, by considering that the critique must firstly focus on the current social framework of scientific production.

The mainstream of quantification in experimental psychology continues to generally use a standardized design, labeled *statistical positivism* (Gigerenzer, 1990b). Quantification requires quantitative measures. Most articles using such measures do so as if these attributes could be measured like the objects studied in physics. Based on these measures, statistical models are used with different problems: (1) confusion between reality, concepts, and variables; (2) errors in the analysis or interpretation of statistical models; and (3) normative vision of the model that neglects singularities and the interdependence of individuals. Criticism of the positivist claims of empirical studies in psychology has been around for a long time. Why does experimental psychology continue to proceed as if this critique did not exist? The fundamental reason is the social function of quantitative psychology. Statistical models allow researchers to publish so-called scientifically valid results (publication bias). Beyond the scientific field, scientific results in psychology contribute in the public space to what Foucault called *bio-power* (Foucault, 1995): the results of experimental psychology not only serve to support public health recommendations but also underpin processes of standardization, control, and regulation.

Because psychology science is different from what it is in natural science (Hacking, 2000), we have to break away from the dominant social practice in psychology. Considering that "Pragmatism starts from the premise that 'thinking is for doing' [. . .] A pragmatist philosophy of science urges scientists to observe what behaviors emerge in the complexity of real life; it encourages active theorizing about individuals' contexts and the way that individuals construe or interpret them" (Gantman et al., 2018, p. 4), measurement and statistical modeling in psychology should be seen as part of a pragmatic approach and not as a protocol proving theoretical hypotheses on individual psychological dynamics.

THE NATURE OF THE PSYCHOLOGICAL ATTRIBUTES AND THE MEASUREMENT ISSUE

The focus of our critique of measurement in psychology is the object being explored by a measurement. Mainstream psychology considers that measuring a mental attribute amounts to considering that a psychological attribute is a true reality, independent of the knower, that can be located physically, in the same way as physics is able to locate its objects. Psychology must break with the dominant epistemology of "biological realism" (Lloyd, 2010; Zachar, 2010). This does not mean that we must return to an instrumentalist/constructivist epistemology. We consider that, in psychology, we need to adopt a pragmatist and realist epistemology (Maul, 2013; Guyon et al., 2018). In psychology, we seek to measure psychological attributes that are real objects but need to be apprehended through the prism of social practice; their ontological nature is different from the objects that physics measures (Searle, 1996). A psychological attribute is an emergent property (Maul, 2013; Held, 2014; Maul et al., 2016; Guyon et al., 2017), the reality of a psychological attribute resides in its functional manifestation (Maul, 2013). A concept in psychology can thus be considered as referring neither to a fixed reality (external to social praxis) nor to a singular construction independent from physical reality (Maul, 2013; Guyon et al., 2017). The categories used in psychology are relational entities, interactive genres (Hacking, 2000). This necessary theory of knowledge for us relates to pragmatism, not in the common sense of the term, but in the acceptation of the philosophy of science (Guyon et al., 2017, 2018; Maul and McGrane, 2017). Pragmatism-realism does not deny the objectivity of knowledge, even if it is a practical objectivity (Maul, 2013; Guyon et al., 2017, 2018). Apprehending reality as being subjectivated is not in contradiction with the consideration that we objectivate reality (Putnam, 1981, 1992), even if the process of objectification is carried out through the prism of tools of representation (language or other).

In consequence, a protocol for validating measurement specific to the field of psychology is therefore needed, breaking with the formal framework of measurement in physics. Such a protocol to validate a measurement in psychology appears to be operational if it is considered as a pragmatic approach (Sherry, 2011; Mari et al., 2012; Guyon et al., 2018; Maul et al., 2018).

STATISTICAL MODELING ISSUE

The term *construct*, since Cronbach and Meehl (1955), has generally been used in psychology to characterize mental attributes in quantitative models. Academic literature points to a confusion on how to apprehend a construct in empirical studies (Slaney, 2001; Borsboom et al., 2009; Lovasz and Slaney, 2013; Maraun and Gabriel, 2013; Markus and Borsboom, 2013; Michell, 2013; Slaney and Racine, 2013). Clear and precise definition of a construct is rare in psychology because of the confusion between concept, variable, and reality (Maraun and Peters, 2005; Maraun and Gabriel, 2013). The statistical model is an abstract and formal representation of associations between variables (mathematical formalism). A clear distinction must be made between the reality (the real psychological attribute), the associated concept (which categories the psychological attribute), and the mathematical formalization of the psychological attribute using a variable. This tension between psychological attribute, concept, and variable generates tensions between substantive theory and statistical model. The statistical model represents the theory in mathematical representation, but there is no equivalence between the two. From the statistical model to the theory, there is the addition of "meaning," that is to say that we move from a mathematical formalism to the substantive theory (Falissard et al., 2013). In addition, when a statistical model is considered validated, there is no statistical method to consider that it correctly models the operationalized objects because of the potential effects of confounding variables or the problem of equivalent models (models with the same statistical validity but with very different theoretical meanings). It is the scientist, in relation with the substantive theory, who will discuss the reasons for considering one model as relevant. Moreover, any normative model neglects singularities and the interdependence of individuals. The empirical regularities detectable by the statistical methods in psychology never constitute knowledge that can be applied to individuals without discussion (Danziger, 1985; Molenaar, 2004; Borsboom et al., 2009; Borsboom and Markus, 2013; Lamiell, 2013, 2019).

The reason why experimental psychology transforms an average model into a valid model for each individual is expressed as follows: "in the thrall of a physical of science and, as a consequence a physical image of man, psychology was forced to eliminate the particular individual" (Gigerenzer, 1990a, p. 29). Statistics in psychology aim to align with the ideals of *determinism* and *objectivity* (Gigerenzer, 1990a). What drives this statistical positivism in psychology stems from three beliefs about quantitative models: (a) that they are intrinsically objective, ensuring objectivity in research work, (b) that they provide estimated values with precision (through fit indicators), and (c) that they ensure scientific rigor (Tafreshi et al., 2016). Therefore, the scientific ideal of physics has become the general ideal and still serves as a reference.

Various articles return to the construction of the quantitative imperative in psychology and the positivism/modernism underpinnings (Danziger, 1985; Gigerenzer, 1990a; Martin, 2003; Michell, 2003a,b). In fact, the results of a psychological statistical model depend on subjective choices: choice of coding of variables, choice of model, and choice of interpretation of results. A psychological statistical model poses a *prototype* (average model) to which no one corresponds and it serves as a kind of *ideal-type* (Niaz, 2005). Moreover, the meaning of a statistical model underlies our philosophical/theoretical beliefs and commitments (Allen and Clough, 2015). We consider that quantitative models are operational in psychology if they are part of a pragmatic approach.

A SOCIAL ISSUE

Danziger (1985) and Porter (1996) showed how quantitative methods were introduced into public life because they were thought to embody the qualities of objectivity and trustworthiness, with an implicit belief in the scientific neutrality of the techniques used. Statistical methodology has become highly institutionalized, providing important criteria for publication policies that became methodological imperatives in the academic literature, even if these methods are erroneous or misapplied (Danziger, 1985; Gigerenzer, 1990a; Lambdin, 2012). There is a rhetoric of scientific language in psychology to foster the authority of knowledge in psychology because it creates the illusion of a scientific validity of results identical to that in the natural sciences (John, 1992).

On the basis of these supposedly objective results, psychological science should be used to help social decisions (Ferguson, 2015). There are many sectors that use such results: tests on children, human resource management, etc. These management tools need to rely on certain results, otherwise their recommendations could appear fragile. Yet, we know that some of the statistical results from psychological quantitative models are wrong; Krantz and Wallsten (2019, p. 132) wrote: "We are horrified by much of the statistical practice in psychology and other research." However, the critical discourse on psychological models is not prominent because these models respond to a social demand. Psychological models are therefore well-embedded in social contexts and issues. The usefulness of psychological models should be understood in relation to the stakeholders' issues, for example, the personality scales widely used by companies while the scientific foundations of which are clearly open to criticism (Lamiell, 2000, 2013; Cramer et al., 2012; Franić et al., 2013). Lacot et al. (2015) rightly say that these psychological models are "primarily ideological." They provide a framework for individuals, where singularity is apprehended only in relation to a norm.

Psychological quantitative models, in the name of objectivity and the determinism of its results, serve as a means of recourse to processes of standardization, control, and regulation, called bio-power by Foucault (1995, 1998). We therefore join Hacking in his critique of the political function of psychology. For Hacking (1998, p. 215), bio-power "engendered the specific technologies of statistics," and bio-power can be extended to "the mind."

CALL FOR A PRAGMATIC APPROACH

We are witnessing a production of psychological models whose objective foundations are sometimes/often open to criticism (Toomela, 2008; Lamiell, 2013) and which can only be understood by their social and political functions in discourse. To be recognized as a "science," the goal of positivist/modernist psychology is to find stable empirical regularities (Tafreshi et al., 2016). When empirically verified, these regularities acquire the status of scientific theories to explain and anticipate the behaviors of individuals (Chirkov and Anderson, 2018). It is indeed the social usage of psychological quantitative models that is criticized, both in the academic field and in the public space. The symbolic weight of statistical results makes it possible to attribute so-called objective facts to individuals, thus making it possible to justify social and political dynamics. Statistical models are more often not actually used to help understand the individual, but they are used for assessment and to set up normative relational frameworks (bio-power). But, this social function of statistical models is not intrinsically linked to the processes of quantitative psychology.

We assert that the scientific approach in psychology must break with the modernist claim. If psychology is a means of intervening in social interaction to support personal approaches (psychiatric pathology, educational assistance, etc.), we must criticize the political function of statistics in psychology. The psychologist must clearly differentiate between singularity and the norm (the average results to which the model refers). More fundamentally, a quantitative model validates an average relational structure between variables (i.e., an abstract codification of real objects) and cannot in itself explain the underlying mechanisms that theory hypothesizes between these real objects (psychological attributes). It must be clearly stated that statistical models can only serve as potential benchmarks, teaching psychologists to distance themselves from these formalizations/representations inscribed in a practice and commitments.

We consider that quantitative models are operational in psychology if they are part of a pragmatic approach: pragmatic approach to conceptualizing psychological attributes, pragmatic approach to measuring psychological attributes, and pragmatic approach to analyzing statistical modeling. We reject the "anything goes" argued by Feyerabend (2008) because there is possible room for quantitative studies in psychology between modernism and post-modernism. By calling on pragmatism and realism, psychology can find the resources to assert itself as a science of the human complex using quantitative studies, breaking with normative practice in academic psychology and normative function in the public space.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

The authors would like to acknowledge Camille Noús, a symbol of scientific collaboration.

REFERENCES

Allen, P. M., and Clough, S. (2015). Philosophical commitments, empirical evidence, and theoretical psychology. *Theory Psychol.* 25, 3–24. doi: 10.1177/0959354314563324

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., and Franic, S. (2009). "The end of construct validity." in *R. W. Lissitz (Ed.), The Concept Of Validity: Revisions, New Directions, And Application. 135–170*. Charlotte, NC: Information Age Publishing.

Borsboom, D., Kievit, R. A., Cervone, D., and Hood, S. B. (2009). "The two disciplines of scientific psychology, or: the disunity of psychology as a working hypothesis," in (eds) J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra, and N. Chaudhary. *Dynamic Process Methodology in the Social and Developmental Sciences* 67–97, (Berlin: Springer). doi: 10.1007/978-0-387-95922-1_4

Borsboom, D., and Markus, K. A. (2013). Truth and evidence in validity theory. *J. Edu. Meas.* 50, 110–114. doi: 10.1111/jedm.12006

Chirkov, V., and Anderson, J. (2018). Statistical positivism versus critical scientific realism. A comparison of two paradigms for motivation research: Part 1. A philosophical and empirical analysis of statistical positivism. *Theory Psychol.* 28, 712–736. doi: 10.1177/0959354318804670

Cramer, A. O. J., Van Der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H. et al. (2012). Dimensions of normal personality as networks in search of equilibrium: you can't like parties if you don't like people. *Eur. J. Personality*, 26, 414–431. doi: 10.1002/per.1866

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957

Danziger, K. (1985). The methodological imperative in psychology. *Philos. Soc. Sci.* 15, 1–13. doi: 10.1177/004839318501500101

Falissard, B., Révah, A., Yang, S., and Fagot-Largeault, A. (2013). The place of words and numbers in psychiatric research. *Philos. Ethics Humanit. Med.* 8:18. doi: 10.1186/1747-5341-8-18

Ferguson, C. J. (2015). "Everybody knows psychology is not a real science": public perceptions of psychology and how we can improve our relationship with policymakers, the scientific community, and the general public. *Am. Psychol.* 70, 527–542. doi: 10.1037/a0039405

Feyerabend, P. K. (2008). *Realism, Rationalism and Scientific Method: Volume 1: Philosophical Papers*. 1st. Edn. Cambridge: Cambridge University Press.

Foucault, M. (1995). *Discipline and Punish: The Birth of the Prison* (2nd ed). New York, NY. Vintage Books.

Foucault, M. (1998). *The History of Sexuality: The Will to Knowledge*. Westminster: Penguin Books.

Franić, S., Borsboom, D., Dolan, C. V., and Boomsma, D. I. (2013). The big five personality traits: psychological entities or statistical constructs?. *Behav. Genet.* 44, 591–604. doi: 10.1007/s10519-013-9625-7

Gantman, A., Gomila, R., Martinez, J. E., Matias, J. N., Paluck, E. L., Starck, J. et al. (2018). A pragmatist philosophy of psychological science and its implications for replication. *Behav. Brain Sci.* 41:e127. doi: 10.1017/S0140525X18000626

Gigerenzer, G. (1990a). *Probabilisitic thinking and the fight against subjectivity. In The Probabilistic Revolution*. (Cambridge, MA: MIT Press). Vol. 2, 11–33.

Gigerenzer, G. (1990b). *The probabilistic revolution in psychology—An overview. In The Probabilistic Revolution*. Cambridge, MA: MIT Press. 7–9.

Guyon, H., Falissard, B., and Kop, J.-L. (2017). Modeling psychological attributes in psychology - an epistemological discussion: network analysis vs. latent variables. *Front. Psychol.* 8:798. doi: 10.3389/fpsyg.2017.00798

Guyon, H., Falissard, B., and Kop, J.-L. (2018). Measurement, ontology and epistemology: psychology needs pragmatism-realism. *Theory Psychol.* 2018, 1–18. doi: 10.1177/0959354318761606

Hacking, I. (1998). *Rewriting the Soul*. Princeton, NJ: Princeton University Press.

Hacking, I. (2000). *The Social Construction of What?*. Cambridge, MA: Harvard University Press. doi: 10.2307/j.ctv1bzfp1z

Held, B. S. (2014). Realism, reification, and monism. *J.Theor. Philos. Psychol.* 34, 187–194. doi: 10.1037/a0035247

John, I. D. (1992). Statistics as rhetoric in psychology. *Aus. Psychol.* 27, 144–149. doi: 10.1080/00050069208257601

Krantz, D. H., and Wallsten, T. S. (2019). Comment on Trendler's (2019) "Conjoint measurement undone". *Theory Psychol.* 29, 129–137. doi: 10.1177/0959354318815767

Lacot, E., Afzali, M. H., and Vautier, S. (2015). Test validation without measurement. *Eur. J. Psychol. Assess.* 32, 1–11. doi: 10.1027/1015-5759/a000253

Lambdin, C. (2012). Significance tests as sorcery: science is empirical—significance tests are not. *Theory Psychol.* 22, 67–90. doi: 10.1177/0959354311429854

Lamiell, J. T. (2000). A periodic table of personality elements? the "Big Five" and trait "psychology" in critical perspective. *J. Theor. Philos. Psychol.* 20, 1–24. doi: 10.1037/h0091211

Lamiell, J. T. (2013). Statisticism in personality psychologists' use of trait constructs: what is it? how was it contracted? is there a cure? *N. Ideas Psychol.* 31, 65–71. doi: 10.1016/j.newideapsych.2011.02.009

Lamiell, J. T. (2019). Re-centering psychology: from variables and statistics to persons and their stories. *Theory Psychol.* 29, 282–284. doi: 10.1177/0959354318766714

Lloyd, D. (2010). Grand challenges in theoretical and philosophical psychology: after psychology? *Front. Psychol.* 1:9. doi: 10.3389/fpsyg.2010.00009

Lovasz, N., and Slaney, K. L. (2013). What makes a hypothetical construct "hypothetical"? Tracing the origins and uses of the 'hypothetical construct' concept in psychological science. *N. Ideas Psychol.* 31, 22–31. doi: 10.1016/j.newideapsych.2011.02.005

Maraun, M. D., and Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *N. Ideas Psychol.* 31, 33–42. doi: 10.1016/j.newideapsych.2011.02.006

Maraun, M. D., and Peters, J. (2005). What does it mean that an issue is conceptual in nature? *J. Personality Assess.* 85, 128–133. doi: 10.1207/s15327752jpa8502_04

Mari, L., Carbone, P., and Petri, D. (2012). Measurement fundamentals: a pragmatic view. *IEEE Transac. Instru. Meas.* 61, 2107–2115. doi: 10.1109/TIM.2012.2193693

Markus, K. A., and Borsboom, D. (2013). *Frontiers of test validity theory: Measurement, causation, and meaning* (2013-15050-000). Milton Park: Routledge Taylor & Francis Group. doi: 10.4324/9780203501207

Martin, J. (2003). Positivism, Quantification and the Phenomena of Psychology. *Theory Psychol.* 13, 33–38. doi: 10.1177/0959354303013001760

Maul, A. (2013). On the ontology of psychological attributes. *Theory Psychol.* 23, 752–769. doi: 10.1177/0959354313506273

Maul, A., Mari, L., Torres Irribarra, D., and Wilson, M. (2018). The quality of measurement results in terms of the structural features of the measurement process. *Measurement* 116, 611–620. doi: 10.1016/j.measurement.2017.08.046

Maul, A., and McGrane, J. (2017). As pragmatic as theft over honest toil: disentangling pragmatism from operationalism. *Measurement Interdiscip. Res. Perspectives*, 15, 2–4. doi: 10.1080/15366367.2017.1342484

Maul, A., Torres Irribarra, D., and Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement* 79, 311–320. doi: 10.1016/j.measurement.2015.11.001

Michell, J. (2003a). Pragmatism, positivism and the quantitative imperative. *Theory Psychol.* 13, 45–52. doi: 10.1177/0959354303013001761

Michell, J. (2003b). The quantitative imperative positivism, naive realism and the place of qualitative methods in psychology. *Theory Psychol.* 13, 5–31. doi: 10.1177/0959354303013001758

Michell, J. (2013). Constructs, inferences, and mental measurement. *N.Ideas Psychol.* 31, 13–21. doi: 10.1016/j.newideapsych.2011.02.004

Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Measurement* 2, 201–218. doi: 10.1207/s15366359mea0204_1

Niaz, M. (2005). The Quantitative Imperative vs the Imperative of Presuppositions. *Theory & Psychology*, 15(2), 247–256. doi: 10.1177/0959354305051367

Porter, T. M. (1996). *Trust in Numbers – The Pursuit of Objectivity in Science and Public Life* (Reprint). Princeton,NJ: Princeton University Press. doi: 10.1515/9781400821617

Putnam, H. (1981). *Reason, Truth and History*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511625398

Putnam, H. (1992). *Realism with a Human Face*. Cambridge, MA: Harvard University Press.

Searle, J. R. (1996). *The Construction of Social Reality*. Westminster: Penguin Books.

Sherry, D. (2011). Thermoscopes, thermometers, and the foundations of measurement. *Stud. Hist. Philos. Sci. Part A* 42, 509–524. doi: 10.1016/j.shpsa.2011.07.001

Slaney, K. L. (2001). On empirical realism and the defining of theoretical terms. *J. Theor. Philos. Psychol.* 21, 132–152. doi: 10.1037/h0091202

Slaney, K., and Racine, T. P. (2013). Constructing an understanding of constructs -. *N. Ideas Psychol.* 31, 1–3. doi: 10.1016/j.newideapsych.2011.02.010

Tafreshi, D., Slaney, K. L., and Neufeld, S. D. (2016). Quantification in psychology: critical analysis of an unreflective practice. *J. Theor. Philos. Psychol.* 36, 233–249. doi: 10.1037/teo0000048

Toomela, A. (2008). Variables in psychology: a critique of quantitative psychology. *Integr. Psychol. Behav. Sci.* 42, 245–265. doi: 10.1007/s12124-008-9059-6

Zachar, P. (2010). The abandonment of latent variables: philosophical considerations. *Behav. Brain Sci.* 33, 177–178. doi: 10.1017/S0140525X10000841

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

TYPE Erratum
PUBLISHED 08 September 2022
DOI 10.3389/fpsyg.2022.1026974

OPEN ACCESS

APPROVED BY
Frontiers Editorial Office,
Frontiers Media SA, Switzerland

*CORRESPONDENCE
Frontiers Production Office
production.office@frontiersin.org

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 24 August 2022
ACCEPTED 24 August 2022
PUBLISHED 08 September 2022

CITATION
Frontiers Production Office (2022) Erratum: Quantification in experimental psychology and pragmatic epistemology: Tension between the scientific imperative and the social imperative. *Front. Psychol.* 13:1026974. doi: 10.3389/fpsyg.2022.1026974

Erratum: Quantification in experimental psychology and pragmatic epistemology: Tension between the scientific imperative and the social imperative

Frontiers Production Office*

Frontiers Media SA, Lausanne, Switzerland

KEYWORDS

bio-power, epistemology, measurement, ontology, pragmatism, quantitative psychology, statistical positivism

An Erratum on

Quantification in experimental psychology and pragmatic epistemology: Tension between the scientific imperative and the social imperative

by Guyon, H., and Nôus, C. (2021). *Front. Psychol.* 11:603617. doi: 10.3389/fpsyg.2020.603617

The name previously presented as the second author is a fictional character, which is contrary to the authorship policies of Frontiers journals.

This name has been removed, and we have placed it in the acknowledgments instead, written as: "The authors would like to acknowledge Camille Noús, a symbol of scientific collaboration."

The publisher apologizes for this mistake. The original version of this article has been updated.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 28 January 2021
doi: 10.3389/fpsyg.2021.609802

Why Psychology Needs to Stop Striving for Novelty and How to Move Towards Theory-Driven Research

Juliane Burghardt[1,2]* *and Alexander Neil Bodansky*[2]

[1] Department of Psychology and Psychodynamics, Karl Landsteiner University of Health Sciences, Krems, Austria,
[2] Department of Social Psychology, Universität Hamburg, Hamburg, Germany

Psychological science is maturing and therefore transitioning from explorative to theory-driven research. While explorative research seeks to find something "new," theory-driven research seeks to elaborate on already known and hence predictable effects. A consequence of these differences is that the quality of explorative and theory-driven research needs to be judged by distinct criterions that optimally support their respective development. Especially, theory-driven research needs to be judged by its methodological rigor. A focus on innovativeness, which is typical for explorative research, will instead incentivize bad research practices (e.g., imprecise theorizing, ignoring previous research, parallel theories). To support the advancement of psychology, we must drop the innovation requirement for theory-driven research and instead require the strongest methods, which are marked by high internal and external validity. Precise theorizing needs to substitute novelty. Theories are advanced by requiring explicit, testable assumptions, and an explicit preference for one theory over another. These explicit and potentially wrong assumptions should not be silenced within the peer-review process, but instead be scrutinized in new publications. Importantly, these changes in scientific conduct need to be supported by senior researchers, especially, in their roles as editors, reviewers, and in the hiring process. An important obstacle to further theory-driven research is to measure scientific merit using researchers' number of publications, which favors theoretically shallow and imprecise writing. Additionally, it makes publications the central target of scientific misconduct even though they are the main source of information for the scientific community and the public. To advance the field, researchers should be judged by their contribution to the scientific community (e.g., exchange with and support of colleagues, and mentoring). Another step to advance psychology is to clearly differentiate between measurement model and theory, and not to overgeneralize based on few stimuli, incidences, or studies. We will use ideas from the theory of science to underline the changes necessary within the field of psychology to overcome this existential replication crisis.

Keywords: explorative research, theory-driven research, innovation, theory of science, replication crisis

OPEN ACCESS

Edited by:
Barbara Hanfstingl,
University of Klagenfurt, Austria

Reviewed by:
Francisco Barceló,
University of the Balearic Islands, Spain
Peter Adriaan Edelsbrunner,
ETH Zürich, Switzerland

***Correspondence:**
Juliane Burghardt
juliane.burghardt@kl.ac.at

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *24 September 2020*
Accepted: *07 January 2021*
Published: *28 January 2021*

Citation:
Burghardt J and Bodansky AN (2021) Why Psychology Needs to Stop Striving for Novelty and How to Move Towards Theory-Driven Research. Front. Psychol. 12:609802. doi: 10.3389/fpsyg.2021.609802

Many factors have contributed to the current replication crisis (see Kerr, 1998; Ioannidis, 2005; Simmons et al., 2011; Casadevall and Fang, 2012; Giner-Sorolla, 2012), which exposed the low replicability of effects in psychological science (Open Science Collaboration, 2015). Replicability is directly tied to the accuracy of measurements (Stanley and Spence, 2014), which is intertwined

with strong theorizing. Both are objectively weaker in psychology than in other disciplines (see Fanelli, 2010). Although frustrating for researchers this, in and of itself, does not imply that psychology is failing. Instead it reflects the age of psychology as a comparatively young discipline in the scientific cannon—while researchers have been studying physics for over 400 years (counting from Galileo Galilei or William Gilbert) experimental psychology has only had 150–180 years to develop (counting from Gustav Fechner or William James). In addition, psychology explores a highly complex research object. Humans are compiled of those very objects physics is studying (i.e., particles or strings); plus, its interactions. It is therefore unsurprising that psychological theories do not *yet* provide strong predictions and constraints (Fanelli, 2010). However, other factors grounded in sociocultural processes do hinder the advancement of psychology. If unchecked they will prevent psychology from becoming as precise as physics. Drawing on the philosophy of science and its insights we will outline five current challenges to psychological research and possible solutions to advance the maturation of psychological science from explorative to theory-driven research.

CHALLENGE ONE: HOW TO PROMOTE RESEARCHERS THAT ADVANCE THE FIELD

The philosophy of science has struggled with the so called "demarcation problem," which defines science and distinguishes it from other human endeavors. This struggle illustrates the difficulty to define the quality of scientific research. To solve this problem, Fleck (1980) argued that research is defined by its fellowship. He described science as a set of social actions, which lead to the development of a collective thought style. A collective thought style implies a certain view of the world (e.g., determining which questions are scientific and worth answering). It defines its own language and appropriate methods to investigate the world in order to gain answers to scientific questions. In an extreme reading of his thoughts, science is nothing more than a social construct. For him, the "truthiness" of scientific facts rests within the breadth and depth of the fellowship of them. Fleck's (1980) constructivist approach to science highlights a specific problem still prevailing in current psychology: Scientific merit cannot be objectively inferred from a theory or a finding. It is thus, very difficult to evaluate the scientific merit of researchers in an impartial and unbiased way.

A socially accepted workaround has become to use the number of publications as a criterion to judge scientific achievement and to evaluate individual researchers. While it is unclear how much faculties and funders actually rely on this criterion to make hiring, promotion, and funding decisions, it is clear that it has become proverbial to publish or perish in order to succeed in science. As a result, researchers strive to increase their publication output.

A presumed advantage of the "number of publications" criterion is its perceived objectivity. However, this objectivity is spurious. Among others, authorship is influenced by social processes and not truly based on the amount or scientific merit of contributions to a publication. For instance, it is arbitrary who receives a co-authorship. Senior scientists may contribute less than junior scientists; sometimes being chief of the department and proofreading the manuscript can suffice to receive a co-authorship as a senior researcher (Stroebe et al., 2012, n. 7). Further, some evidence suggests that gender influences publication outcomes. One study found that female Ph.D. students are less likely to author papers than male Ph.D. students, even though they put in more time (Feldon et al., 2017). Papers with female first authors are reviewed longer and more critically than those of male first authors arguably because they are held to higher standards (Hengel, 2017; Fox and Paine, 2019), which would make it more time-consuming for women to publish.

Importantly, the number of publications does not consider the quality of articles. Number of publications rewards publishing many, potentially theoretically shallow articles. As an unwanted consequence psychological literature is inflated by many "parallel" theories (Glöckner and Betsch, 2011), a plethora of "sexy" singular effects (Fiedler, 2017), different "mini-theories" (Glöckner and Betsch, 2011), or simple analogies, which suffer from low degrees of precision and universality[1]. The lack of precision in theory building frequently eliminated the possibility to test theories against each other (Glöckner and Betsch, 2011) because they did not contain enough assumptions, thus, allowing multiple theories to coexist. Psychological effects were often reported outside of established theoretical structures, thus, ignoring existing theories and undermining the integration of knowledge into an overarching theoretical understanding. Variables highly similar to previously existing constructs were often introduced without referring to the related construct ("déjà-variable"; Hagger, 2014). We believe that the reason for this phenomenon, is often not a lack of knowledge, but the understanding that avoiding controversial claims by producing theoretically vague writing will increase the publication chances and facilitate the review process because it invites less scrutiny by critical reviews. To increase output and withstand the pressure to "publish or perish" highly similar articles were produced or series of studies were divided into multiple articles (i.e., salami publishing).

Thus, using number of publications as a measure for scientific merit is problematic, in and of itself. In addition, it produces high long-term costs for the whole field. The inflated literature, produced to be hired, increases the amount scientists need to read to identify new information. Furthermore, scientists spend more time reviewing articles of other researchers (Casadevall and Fang, 2012). Additionally, the need to publish for one's own career leads to a rat race, which does not reward rigorous and hence time-consuming research. This rat race is won by fast and effective publishing and foreseeing reviewers' reactions to one's own submissions. Hence, authors write to please and convince the gatekeepers within the publication system. This discourages from reporting ambiguities within research, thereby leading to tactical omissions and hence, toward the behavior underlying the

[1]Precision describes the number of possible implementations allowed by predictions. Universality describes the number of situations that the theory can be applied to Popper (2002).

replication crisis. Further, the focus on publications for evaluating individuals will make these publications the main target of scientific misconduct and other manipulations, even though these publications are the main source of scientific knowledge for the general public.

Thus, incentives within psychology have created strong conflicts between behavior that advances science and behavior that advances the career of individual researchers (Casadevall and Fang, 2012; Giner-Sorolla, 2012; Nosek et al., 2012; Pickett, 2017). An optimal scientific environment needs to create as much overlap between what is good for science and the scientist as possible (see Pickett, 2017). To reduce this conflict, promotion and hiring criterions need to be based on behavior that supports the integrity of science in the long run. As a first step, publications need to become more valid measures of scientific merit. Thus, contributions of individual authors need to be identifiable (Casadevall and Fang, 2012). Of course, this can only be a makeshift solution. We are convinced that number of publications should to be dropped as a hiring or evaluation criterion in favor of in-depth evaluations of the scientific quality of publications (Casadevall and Fang, 2012). Some institutions have already adjusted their procedures accordingly and invite candidates to submit a limited list of articles. For instance, the German Research Foundation limits the number of publications per applicant to ten (Deutsche Forschungsgemeinschaft, 2020). Importantly, if institutions and senior researchers choose to favor quality of research over quantity, this must be effectively and repeatedly communicated to create a norm within the field. We recommend faculties that support a quality criterion to commit to this in job postings. Even faculties that never relied on the number of publications should make this public to support the quality orientation within the field.

Instead of focusing on publications, Pickett (2017) suggested promoting researchers that have positively impacted the scientific system. Excellent researchers increase the influx of talented new researchers, support the productivity of other scientists, and make the field a better and more productive place. This stands in contrast to researchers who hoard resources, practice favoritism, or generally conduct their projects in a competitive rather than a collaborative manner (see Anderson et al., 2008). Researchers with positive impact on the field should regularly assist colleagues by sharing ideas, materials, programs, and advice. Further, they should have fostered mentees, especially, diverse mentees are a strong indicator of excellent collaboration and leadership skills. To measure this, it is possible to use websites like google scholar or ResearchGate, which offer lists of frequent co-authors. Lists with diverse co-author's (e.g., regarding gender, nationality, or ethnicity) are a sign of someone with a history of working with people form many backgrounds. Promoting researchers who support scientists from diverse backgrounds is not simply an idle moral goal; diversity is crucial for knowledge accumulation. Excluding input from researchers from specific groups (e.g., first generation academics, women) results in a tremendous loss of know-how and expertise for science (Merton, 1979). As science is quintessentially a collaborative endeavor, individuals that rely solely on the expertise of a small group of socially homogenous fellow researchers will inevitably constrain the potential of the field. There is ample evidence for the increased productivity of diverse teams in experimental and field studies in a wide variety of tasks (Roberson et al., 2013). Importantly, the behavior toward colleagues and mentees from minority or minoritized groups is a potent indicator of problematic interpersonal behavior (i.e., bullying or discrimination). For instance, bullies prefer victims with less power, such as subordinates and members of disadvantaged groups (Salin, 2003). Thus, having mentored individuals with diverse background is a litmus test of an excellent leader, which is what a principal investigator should be. As a means to this end, some universities let undergraduate students evaluate job candidates for faculty positions. Undergraduates are more likely to trigger problematic interpersonal behavior. Requiring principle investigators to have a history of supporting others, especially subordinates, benefits both the scientific productivity as well as the working atmosphere. Thus, scientific merit should be evaluated by criterions that have a long-lasting positive effect on the academic system and not by criterions, which lead to negative consequences. Fleck (1980) conceptualized science as a social endeavor. It should be as inclusive as possible and not a race of individuals to be the single most prolific publisher.

CHALLENGE TWO: DEFINING WHICH RESEARCH SHOULD BE PUBLISHED OR FUNDED—THE AGE OF DISCOVERY IS OVER, NOW IT'S TIME FOR PUZZLE-SOLVING

Even though our recommendations suggest to devalue the number of publication criterion, publishing one's findings is essential to science. The criteria used to evaluate manuscripts within the peer-review process are a crucial factor that shape the scientific conduct. We will draw on Kuhn's (1970, 1996) insights into the theory of science to scrutinize the review criteria. Kuhn describes that science evolves through a progression of different phases. Early stage science, termed protoscience, develops into paradigmatic science[2]. Protoscience (still) lacks broad theories and thus, often relies on explorative as opposed to theory-driven research. It therefore explores ideas in a relatively random fashion and uses idiosyncratic methods. We argue that psychology is in the process of developing from protoscience into paradigmatic science. For this transition, the field has to adopt and consent to a paradigm. A research paradigm defines relevant theories, (measurement) instruments, values and metaphysical assumptions, which are kept relatively constant. To succeed, each research community needs to commit to a paradigm, to meaningfully communicate with each other and to explain as many findings as possible in a way that can be shared, taught, and utilized. Explorative research has to cease when the scientific community consents on a paradigm.

[2]Kuhn referred to paradigmatic science as "normal science." As a trained physicist his focus lied on physics, which he equated with "normal science." Since this terminology is biased, we instead use the term paradigmatic science.

Kuhn (1996) described research conducted within paradigmatic science as "solving puzzles," meaning, that researchers have a rather clear understanding of what they expect to find prior to conducting their studies and of the tools needed to show the predicted effects. They need to fill the gap with the missing piece; the solution of the overall puzzle is well-defined. A research finding may lead to changes in the details of the auxiliary assumptions or measurement models (see below). However, the paradigm as well as the associated theory should remain unchanged.

Only when a critical number of anomalies and unexplained findings accumulate will researchers begin to question their paradigm. In the subsequent "revolutionary phase" a new scientific paradigm is developed. These phases of revolution, however, are rare and do not constitute everyday research practice. Thus, revolutionary science is unusual and more importantly cannot be planned ahead, for instance, within tenure-track or a grant proposal. Hence, scientists should not strive for it. Instead, they should strive to conduct research in the spirit of a Kuhnian paradigmatic phase, which is marked by high precision and rigorous application of sound methods—not by breakthroughs. This research emphasizes "dull" routine jobs like theory improvement or increasing the precision of measurements as the essence of good (paradigmatic) research.

As a consequence, new criteria for evaluating submissions or grant proposals that promote good paradigmatic research need to be established. If we follow Kuhn's logic and we believe that psychology needs to transition into a paradigmatic science, then "innovativeness" can no longer be used as an adequate criterion to judge research, because it counteracts the objectives of paradigmatic research. Currently, "innovativeness" is frequently used as a central criterion to evaluate the contribution of an article or research project (Giner-Sorolla, 2012; Nosek et al., 2012). Innovation is equated with "novelty," implying that an unknown effect or a new theory is preferred. Thereby, the criterion of innovation actively disincentives paradigmatic science: Studies not seen as novel enough can easily be dismissed as "trivial." Results are frequently rejected by journals because they are "merely incremental" (Giner-Sorolla, 2012; Organization Science, 2020), which creates incentives that hinder good paradigmatic research. For instance, repeating a study on stereotypes with the gender category instead of the race category is an ideal study in paradigmatic science, because it tests whether the existing assumptions hold for new stimuli. Testing whether previous assumptions hold for other stimuli is furthermore also a test about the ceteris paribus conditions (when all other things are equal) of a theory; meaning that results replicate. It is thus, good paradigmatic research. These "dull" research programs truly add to the accumulation of robust knowledge instead of piling up fancy, new looking fast fashion research.

Furthermore, the focus on innovation has the tendency to undervalue research that draws attention to old but ongoing problems for which a solution has already been suggested (e.g., Clark, 1973; Judd et al., 2012). The community should not assume that a point is made successfully just because someone already made it. Currently such valuable "reminders" are either unpublishable or relegated to less prestigious outlets. This tendency understates their scientific contribution and prevents arguments from informing future research. The emphasis on novelty remains unchallenged even though recent replication attempts have highlighted the sophistication involved in a successful replication of a previous study (e.g., Maxwell et al., 2015; Noah et al., 2018; Bressan, 2019).

To overcome these problems, we advocate to redefine the value of scientific contributions, especially, to stop valuing innovation and novelty. Following Kuhn's observations about paradigmatic science, unpredicted results are more often than not a sign of imprecise theories. They are thus, not something researchers should strive for. As shown above, rewarding innovation in science has the tendency to hinder incremental work. Researchers should systematically test predictions made by their theory. In contrast, a focus on innovativeness will have multiple negative effect: First, innovativeness is often much more difficult to proof than accuracy. Accuracy can, at least partly, be quantified in terms of reliability and validity; thus, offering relatively objective measures. Therefore, judging the accuracy of a study should show high interrater (i.e., reviewer) consensus (e.g., a study may explain 10% more variance in a given design than previous studies). However, innovativeness is much more difficult to recognize let alone to quantify. Innovativeness is often identified by an absence of a similar study, theory, or effect. This absence is, however, only recognized by a reviewer that has a perfectly fitting expertise. Something that is currently difficult to ensure but will become increasingly difficult due to the increased differentiation within the field. Thus, the nature of the innovativeness criterion makes it much easier to cheat by omitting relevant work and much harder for reviewers to recognize. How incapable we are to identify good work when seeking innovation becomes clear from the fact that work that later Nobel prizes were based on, has consistently been published in less prestigious journals (Kumar, 2009).

Second, a logical outcome of preferring novelty over accuracy is that methodologically more advanced follow-up studies are frequently published in lower ranking journals than methodologically less advanced and less accurate first demonstrations (e.g., Sherman and Bessenoff, 1999; Sherman et al., 2003), which undermines their value. The easiest remedy for this is to use an ongoing review system for articles (Nosek and Bar-Anan, 2012). This would imply that, similar to online rating engines, publications could constantly be (re)rated according to their methodological rigor and impact on the field. This would assure that over time the methodologically more rigorous study would have a better rating than the earlier, less precise versions and thereby, would receive more attention from researchers. This solution is technically easy to implement. A more comprehensive solution would imply to abandon the classical journal system and instead publish all articles in the same database. Journals would merely "advertise" articles, which is already successfully implemented in the field of physics (Nosek and Bar-Anan, 2012).

Third, the negative effects of focusing on innovativeness can even reach beyond the academic system when researchers decide to disseminate early findings without replications. For instance, the effect presented in the most watched TED-talk (i.e., power posing) is now challenged (Jonas et al., 2017). This is unsurprising

since the corresponding studies were published only 2 years prior to the talk (Carney et al., 2010). Such fast communication of unreplicated results bears the risk to delegitimize science and its conduct in the eye of the general public.

Fourth, the difficulty to measure innovativeness introduces additional ambiguity into the review process. Ambiguity is known to increase the impact of stereotypes (e.g., Norton et al., 2004) and as a consequence racism, sexism, ageism and other biases such as motivated reasoning steaming from conflict of interests. Since innovativeness is associated with scientific merit it is part of the ideal that individuals need to fulfill to become researchers. Such norms will likely influence hiring and career decisions via other- and self-selection (Eagly and Karau, 2002; Heilman, 2012). It is thus noteworthy, that women are less likely to describe their research as "new" or "innovative" than men (Lerchenmueller et al., 2019). Therefore, the innovativeness criterion may elevate barriers women perceive when entering or pursuing a career in science.

Fifth, a caveat of innovation is its association to creativity. Innovation and creativity are linked to dishonesty (Khessina et al., 2018). Both, a creative disposition and a creative mindset were shown to increase cheating behavior (Gino and Ariely, 2012; Gino and Wiltermuth, 2014). This is presumably because seeking creativity encourages actors to ignore rules and thereby provides a justification for unethical behavior. Giner-Sorolla (2012) argued that creativity is linked to an artistic conceptualization of science in which output has to be aesthetically pleasing and therefore narratives and data have to be "beautiful." This contradicts the reality of empirical research and therefore potentially motivates problematic conduct.

Sixth, relying on accuracy instead of innovation might even reduce the negative effects of "scooping." Scooping, in its most extreme form, implies that a researcher publishes a study that they copied from another researcher, for instance from a conference presentation. As a result, the person who originally had the idea will find it difficult to publish it. If being the first was valued less, it should become more rewarding to advance ideas than to simply copy them. As a result, both studies would be publishable. It might even become more beneficial for both researchers to collaborate than to compete.

Thus, the development into a paradigmatic science makes innovativeness an obsolete criterion to judge publications, grants, and job candidates on. Psychological science should instead emphasize the value of accuracy. Accuracy by its nature, needs to be evaluated continuously. Even though Kuhn (1970) explicitly refrained from providing advice about how to arrive at the paradigmatic stage, it is clear that it requires both broad and precise theories. Agreeing upon such an overarching theory must be a crucial goal in psychology. Some broad models already exist; however, they are not widely adopted. A successful model needs to integrate many different assumptions for instance, about psychological constructs and about the brain's architecture. The Hierarchically Mechanistic Mind (HMM) theory offers such a broad framework[3]. Grounded in evolutionary psychology, it combines assumptions from psychology and neuroscience to outline perception, cognition, and behavior based on a hierarchical, dynamic brain architecture. Such a highly universal model, would need to be integrated with models with higher precision in specific areas. For instance, the PSI theory (Kazén and Quirin, 2018), which integrates assumptions about motivational and volitional processes to explain personality. In the resulting universal and precise framework, future researchers would fill in well-defined gaps in knowledge with the matching "puzzle piece."

CHALLENGE THREE: HOW TO CONDUCT GOOD RESEARCH

If Kuhnian "puzzles" ought to be solved, then there must be a scientific methodology to solve them. Popper argued that the scientific process implies that researchers choose a theory, deduce hypothesis from it and test them empirically. If the data doesn't support the hypothesis, the theory should be given up. This idea was revised and refined by Lakatos (1970), who argued that in order to derive a research hypothesis from a theory, researchers need to make additional auxiliary hypotheses and assumptions. Further, specific measurement methods and theorizing about these measures (measurement models) are needed to gather data. Finally, all hypothesis can only be tested under ceteris paribus conditions (when all other things are equal, and no intervening factors present). Research following Lakatos has to reconcile all these components to test the proposed hypothesis. If the empirical data does not fit the stated research hypothesis, it is not possible to identify the source of this imbalance due to the many factors involved in the process. The reason for the data not matching the hypothesis could lie within the theory, the auxiliary hypothesis and assumption, the stated hypothesis, the measurement, the theory of measurement, or not fulfilling the ceteris paribus conditions. The same can also be true, if the data *does* match the proposed research hypothesis. Since, this could also result form an error within the research process; for example, a confounded measurement. Therefore, data alone can neither verify nor falsify a theory. Rather, data can initiate a reconsideration of the whole research process and its parts. This requires fine tuning and adaptations in every part of the research process and repeated tests of scientific assumptions against new data. Thus, scientists have to balance all elements of the research process (theory, auxiliary hypothesis, hypothesis, measurement, ceteris paribus and data), while not knowing, which of these elements is ultimately false, and which can be relied on until it also has been proven wrong. Lakatos (1970) compared the research process to drawing piles into a swamp to erect a building on top of them, not knowing which pile would last for the time being.

Lakatos' (1970) conceptualization of science helps to outline ideal theory-driven research. The implications especially pertain to the publication process. Since the replication crisis uncovered many gaps in our knowledge, some scientists may respond by displaying more restrain in postulating assumptions and therefore make fewer predictions than previously. In contrast, Popper and Lakatos urge individual researchers to make clear

[3]We thank an anonymous reviewer for this suggestion.

predictions and claims. The goal they propose is not to be proven correct—quite the contrary, theories are by default incorrect because they oversimplify—instead, making clear predictions and be proven wrong is the ideal way to advance theory development and hence understanding. This misconception frequently surfaces, when theories are tested against each other by researchers who claim to be impartial about their preferred theory. This impartiality can be motivated by trying to sell a null effect finding or to assure *a priori* that any outcome of an experiment could be valid, guaranteeing a publication. According to Lakatos (1970), however, testing theories against each other necessitates the opposite of impartiality. Researches ought to take a stand on which theory is better and why. They need to show a "dare-devil attitude" (Lakatos, 1970, p. 112), not an impartial one and hence should be motived to prove that their theory is better. In extreme cases of impartiality, researches "pick" different theories for each publication based on their results, while never accumulating the information they found into a stance that could be proven wrong. They thereby circumvent the burden of accountability. Accumulating restraining assumptions is crucial for theory development and hence, necessary for scientific progress. Establishing boundaries of one's own theory should be a success and not a stain. To overcome these obstacles we reaffirm the suggestions made by Glöckner and Betsch (2011): Scientists need to (1) make strong claims ("state a finite set of definitions and propositions that together constitute their theory" p. 717), (2) "formulate the propositions in such a way that the theory as a whole clearly predicts particular states of the world to occur and others not to occur." (p. 717), and (3) "The authors should be obliged to explain which kind of empirical observations they would consider a fundamental violation of their theory" (p. 717).

The reluctance to take a strong stand about theories and predictions seems to be based on the expectation that these preferences are often attacked within the review process. This is a crucial obstacle for psychology's transition into a paradigmatic science. Theoretical assumptions and individual preferences should not be buried in the review process. In a beneficial review process reviewers and editors have to assure the logical clarity of the theory, hypothesis, auxiliary assumptions, measurement, and ceteris paribus conditions. However, personal preferences for one theory or hypothesis over another should never influence the course or outcome of a review. Any in principle reasonable (i.e., logically consistent and empirically not repeatedly disproven) assumption is a valid contribution to a theory, irrespective of one's own preferences. The review process is not the right outlet to criticize assumptions that seem at odds with reviewers' or editor's own views. Reviewers and editors should relegate such criticism to a new publication or comment.

However, for the field to change successfully, it is not sufficient to act in line with these suggestions. Importantly, senior researchers should explicitly commit to them by welcoming reasonable hypotheses, as well as supporting explicit assumptions and constrains within the review process (maybe sport a "theory support" batch). No author will follow these guidelines until a critical mass of senior researchers (i.e., editors, reviewers, and supervisors) explicitly supports them.

The main task of reviewers and editors should be to assure the logical clarity of the theory, hypothesis, auxiliary hypothesis and assumptions, measurement, and ceteris paribus conditions. Contradictions between assumptions within a submission need to be resolved. Predictions that conflict with previous findings need to be discussed. They should only be changed if there is an overwhelming number of evidences against them. In this case the article should argue about moderating conditions or context influences that explain why authors assume that a prediction that ostensibly contradicts previous findings is valid. Importantly, research should not be solely judged by its findings, rather the research process as a whole should be evaluated. As Lakatos (1970) pointed out, science is not about findings or data but the interplay of data with theories, auxiliary assumptions, measurement methods, and ceteris paribus. Currently, many reviews put too little emphasis on the measurement models and their adequacy. To evaluate the appropriateness of the review process and to prevent biases, the entire review process should be published alongside the article to allow other scientists to evaluate the process (Nosek and Bar-Anan, 2012; Wicherts et al., 2012).

Lakatos' (1970) ideas define good research as theory-driven. However, based on psychology's short history, it occasionally cannot provide a theory or lacks a sufficiently precise estimate of an effect. Without a scientific paradigm, a young discipline needs to establish effects in an explorative or descriptive manner. It is not useful to force authors to retroactively apply a theory to their findings if the study was not conducted based on said theory or was data-driven. The quality of this descriptive research should be judged based on its methodological rigor (great recent examples provide Smith and Hofmann, 2016; Zwebner et al., 2017; Ray et al., 2019). Better descriptive research has higher internal and external validity. It should create assessments closer to reality, use more accurate dependent measures, bigger and more representative stimulus and participant samples, or non-reactive measures. However, once a theory is established it should be used until it is refuted, or a better theory is proposed. Better theories are marked by higher precision or universality (Popper, 2002). They include more constrains and hence, are more likely to be falsified. Additionally, a new theory that repeats old predictions and constrains but puts them into a broader context is an important improvement because it increases universality. Small theories or singular findings without theoretical foundations should increasingly disappear even, if they are "sexy" or surprising (Fiedler, 2017).

Even though, theory-driven research must be the goal, it is in general, unproblematic to engage in explorative research. The main risk for science steams from mixing theory-driven with explorative approaches. A highly popular chapter on academic writing illustrated such a mix (Bem, 2002). The chapter encouraged researchers to analyze data in any way possible—which is adequate in explorative research—however it also advised scientists to write the results as if they had been predicted, irrespective of the original predictions. Thus, the explorative analysis was followed by theory-driven elaborations; explorative and theory-driven research were inadequately mixed. Where explorative research is sold as theory-driven, the data is used twice: First, to discover a new hypothesis and second, to test

that hypothesis (Kerr, 1998; Wagenmakers et al., 2011). Thus, researchers skip the necessary replication. Data is no longer used to challenge the theory, auxiliary assumptions, or measurements, instead the theory is picked in response to the data.

In conclusion, scientists in psychology currently conduct two different types of research: theory-driven and explorative research. Until psychology adopts overarching paradigms, it seems useful to preserve both types of research since explorative research can be fast to provide new impulses. They should, however, be clearly separated from theory-driven research. The most straightforward solution are separate journal sections reserved for explorative/descriptive and theory-driven research. These sections should apply different evaluation criteria and approaches. Contributions for theory-driven sections would be judged by their increase in accuracy in theory and measurement (see also Wagenmakers et al., 2011), while explorative findings could be judged by their internal and external validity. The differentiation into sections would support the implementation of different evaluation criteria and create a demand for theory-driven research. Thus, it would incentivize researchers to conduct theory-driven research and also elevate the value of theory-driven research. The long-term goal for psychology would remain to transition into theory-based research.

CHALLENGE FOUR: HARMONIZING THEORY AND EVIDENCE

Holzkamp (1981) distinguished between theoretical sentences and empirical sentences: Theoretical sentences are generalized ideas (e.g., stereotypes influence behavior), while empirical sentences are statements about specific observations (e.g., Jamal is described as more threatening than David). A subtype of empirical sentences are experimental sentences that pertain to specific observations within experimental settings in contrast to naturalistic settings. Holzkamp (1981) argued that experimental sentences can be understood as "now-and-here-data" meaning, that they are the result of the given experimental context (e.g., room, demand effects, stimuli, experimental method, experimenter). His work emphasized the gap between experimental sentences and empirical sentences: Experiments are artificial products and can only try to approximate the empirical reality outside of the laboratory. The gap between experimental sentences and empirical sentences is rather large within the field of psychology as the laboratory does not resemble everyday life. Instead, it is a highly artificial environment in which people act as test subjects. This limits their ability to act as independent agents, which renders their behavior irrelevant to everyday life. This gap is problematic, but theoretically it could be bridged by clever experimental designs. There is however a second gap, between empirical and theoretical sentences, which is even more relevant. Theoretical sentences always have multitudes of possible meanings, while each empirical sentence can only represent one of these meanings. But, if there are multiple meanings to each theoretical sentence, then any empirical sentence is never equivalent with the complete meaning of any theoretical sentence. Holzkamp (1981) concluded that no "here-and-now-data" can ever verify or even falsify theoretical sentences. Thus, measurements should never be equated to assumed theoretical constructs.

This insight is crucial for the advancement of psychology: The benefits of theory-driven research might tempt some researches to infuse all levels of analysis with theoretical meaning. However, theories need to be strictly separated from measurement models and thus concrete observations. For instance, behavioral observations need to be separated from the mental processes that are assumed to cause them (see De Houwer et al., 2013). Measurement methods need to be separated from the presumed underlying theoretical constructs and processes (Sherman and Klein, in press). It would be wrong to jump to the conclusion that using an "implicit measure" amounts to having measured only "automatic" processes (e.g., Calanchini et al., 2014). Put differently, psychological research should refrain from over-theorizing empirical data; that is, equating theory and measurement model.

Holzkamp's (1981) distinctions highlight the necessity to bridge the gaps between empirical and theoretical sentences in the most consistent, reliable, and transparent way. This highlights the crucial need for replications in different contexts, under different circumstances, with different stimuli and participants. These replications generate a plethora of empirical sentences that potentially are instances of theoretical sentences. However, for this to succeed, theories need to define as many auxiliary conditions for an effect to occur as possible, as this will increase the chance fur successful replication. For instance, Noah et al. (2018) suggested that a previous high-profile replication effort of the facial-feedback effect had failed because it introduced video recordings to the original paradigm. The replication seemed to have failed because the feeling of being observed blocked the predicted facial-feedback effect. Thus, introducing this assumption into the theory will increase replication chances. Well defined theories, with more specifications about context and moderators will allow a more thorough evaluation of the success or failure of a replication. For instance, Bem (2011) suggested that people can respond to stimuli that will be presented to them in the future (precognition). Within his set of studies, the effect was found for erotic stimuli and sometimes for neutral stimuli, but sometimes it was absent for neutral stimuli. The effect on neutral stimuli could be interpreted as an unsuccessful replication. However, no rational was provided as to when valence should or should not affect precognition (Bem, 2011; Rouder and Morey, 2011). Therefore, Bem (2011) argued that the replication was successful, while Rouder and Morey (2011) argued that it was unsuccessful. These conflicting interpretations highlight the necessity to clarify circumstances, moderators and ceteris paribus for every theory as extensively as possible, so that replications and their success are less open for interpretation.

CHALLENGE FIVE: ESTABLISHING A CHECKLIST FOR GOOD RESEARCH

Latour and Woolgar (1986) described how statements develop into consented upon scientific facts. According to their analysis, science is based on "literary inscriptions." This means that scientists use numbers and words as placeholder to study natural

phenomenon by ascribing digits or labels to these natural entities. Without assigning numbers or language to researched phenomenon (e.g., coding if a participant is a man or a woman) it is impossible to systematically study anything and derive conclusions about recurring trends. Therefore, literary inscription is the first step to construct order in an, in other ways, chaotic system. Latour and Woolgar (1986) pointed out that scientific facts are the product of "sorting, picking up and enclosing" (1986, p. 247) of the inscriptions given to the studied material. Sorting entails which studied material is inscripted and how. Picking up entails the conscious decision to look at certain inscriptions (e.g., man and women) and not at others (e.g., non-binary individuals) and enclosing entails the act of integrating inscriptions in a way that an effect or scientific fact can be seen, for instance, within a statistical diagram.

> "The whole series of transformations, between the rats from which samples are initially extracted and the curve which finally appears in publication, involves an enormous quantity of sophisticated apparatus (. . .). By contrast with the expense and bulk of this apparatus, the end product is no more than a curve, a diagram, or a table of figures written on a frail sheet of paper. It is this document however, which is scrutinized by participants for its "significance" and which is used as "evidence" in part of an argument or in an article"
>
> (Latour and Woolgar, 1986, p. 50).

The sophisticated apparatus, which is necessary to sort random noise from relevant data and to gain agreed on facts is cost intensive. From Latour's and Woolgar's perspective scientific facts are not so much found but manmade things. They are constructed under rules. Science is seen as the human striving to organize the world by choosing to look at "what" and "under which conditions" in order to mold an entity into a reliable, agreed on phenomenon. The only thing hindering the production of infinite scientific facts are the costs associated with the production of facts. Therefore, science is seen as an ongoing project that seeks the most cost-effective way to organize the world. This view on science highlights the importance of the rules that guide the process to establish agreed upon scientific facts and leads us to propose 10 procedures, which should be adhered to.

First, the methods of data analysis should be defined prior to the analysis (Simmons et al., 2011; Wagenmakers et al., 2011). Preregistrations seem to be an ideal tool to force researches to take a stand. However, we don't believe that this method has lived up to its promise yet. Scientists still need to appreciate preregistrations as an opportunity to make theoretical predictions and to protect themselves from self-serving biases (Nuzzo, 2015; Nosek et al., 2018).

Second, stimulus materials, computer code, and raw data should be publicly available (e.g., Wagenmakers et al., 2011; Giner-Sorolla, 2012) to allow for independent checks and further analysis.

Third, internal validity is increased by rigorous control of stimuli, which results from thorough pretests. Consequently, researchers need to evaluate and reduce confounds within stimuli. Subsequently, stimuli should either be representative or selected to be extreme on a criterion defined by the theory. They should also be made freely available to allow for *post hoc* in-depth analysis of the stimuli and to ensure easy replications.

Fourth, external validity should be increased by more naturalistic stimuli, which should be representative and span all possible categories (Brunswik, 1955, 1956). As a first step including many stimuli will be helpful, however, ideally they should represent the nature and distribution of stimuli in reality (Fiedler, 2000). Further, study settings should be naturalistic (Holzkamp, 1981).

Fifth, participants should represent many ages, sexes, education levels, or cultures. Especially, they should be relevant to the research question at hand (Henrich et al., 2010; Landers and Behrend, 2015). For instance, studies on hiring decisions should rely on raters with hiring experience.

Sixth, experimenter bias should be controlled equally strongly as ceiling effects and experimental mortality (i.e., selective dropout of participants), as well as any effects of one measure on the following measure.

Seventh, dependent variables should be representative and relevant. For instance, hiring decisions are categorical decisions, nevertheless, research commonly uses continuous measures (e.g., Uhlmann and Cohen, 2005).

Eighth, it is necessary to rule out reactive or interaction effects of testing where a pretest increases the scores on a posttest. This can be achieved within the Solomon four group design.

Ninth, effects should be replicable in multiple measures as not to fall for one measurement error.

Tenth, responses to the replication crisis have encouraged more rigorous statistical methods especially, requests to reduce the alpha-error (e.g., Nosek et al., 2018). The Bayesian approach constitutes a helpful alternative to frequentist *p*-values (Kruschke, 2014; Rouder, 2015; Dettweiler, 2019). This approach explicitly differentiates between theory and measurement models. The priors force researchers to state constrains, which is in line with our understanding of science as proposed above. The priors reinforce knowledge accumulation. Bayes allows to test null-hypothesis and benefits more from increased measurement accuracy than frequentist models that rely on *p*-values only.

However, any statistic will fail if it is applied to inapt theories. Even the most advanced statistics can't safeguard against a system that rewards those scientists who were lucky "to find" something new. Therefore, as stated above, new publication guidelines are necessary to encourage sustainable research.

The previous sections have outlined many epistemological ideas about how science should be conducted. Previous thinkers have pointed out that scientific merit is socially constructed. They described that theory-driven research leads to better scientific outcomes when researchers are able to carefully balance theory, auxiliary assumptions, measurement models, ceteris paribus conditions, and data. Researchers should always be aware of the context and conditions they create within an experimental paradigm and how it impacts participants' responses. Via successive sorting and picking of relevant information, the community can, however, identify the current consent within a field.

On the one hand, it seems these ideas are well established and rarely refuted, on the other hand, the scientific conduct

in psychological science often deviates from these ideals. The reason for this is that the incentives within psychological science often do not support best practices. To fulfill the potential to become a fully realized paradigmatic science psychology will need to change the incentive structures to align the behaviors that are good for the individual scientist and the scientific community. We need to support the use of broad and comprehensive theories with a consistent emphasis on measurement models. The goal must be to adopt an overarching research paradigm, that includes as many aspects of human life as possible accompanied by measures for these factors. We need to request replications that generalize our findings to new stimuli, individuals, measurements, and contexts, overcoming the popularity of innovation. We need to use the review process as a means to strengthen our ideas and not to protect our theories or censor ideas that contradict our intuition. When psychology can offer broad theories over its entire area it will eventually see more successful replications and will grow as a field.

AUTHOR CONTRIBUTIONS

JB created the first draft. AB substantially rewrote it. Both revised the resulting draft and finalized it.

FUNDING

The work was made possible by a fellowship of the Alexander-von-Humboldt-Foundation awarded to JB. Further it was supported by the Open Access Publishing Fund of the Karl Landsteiner University of Health Sciences, Krems, Austria.

REFERENCES

Anderson, M., Ronning, E., De Vries, R., and Martinson, B. (2008). The perverse effect of competition on scientists', work and relationships. *Sci. Eng. Ethics* 13, 437–461. doi: 10.1007/s11948-007-9042-5

Bem, D. J. (2002). "Writing the empirical journal article," in *The Compleat Academic: A Career Guide*, eds J. M. Darley, M. P. Zanna, and H. L. Roediger, III (Washington, DC: American Psychological Association).

Bem, D. J. (2011). Feeling the future: experimental evidence for anomalous retroactive influences on cognition and affect. *J. Pers. Soc. Psychol.* 100, 407–425. doi: 10.1037/a0021524

Bressan, P. (2019). Confounds in "Failed" replications. *Front. Psychol.* 10:1884. doi: 10.3389/fpsyg.2019.01884

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychol. Rev.* 62, 193–217. doi: 10.1037/h0047470

Brunswik, E. (1956). *Perception and the Representative Design of Psychological Experiments*. Berkeley, CA: University of California Press.

Calanchini, J., Sherman, J. W., Klauer, K. C., and Lai, C. K. (2014). Attitudinal and non-attitudinal components of IAT performance. *Pers. Soc. Psychol. Bull.* 40, 1285–1296. doi: 10.1177/0146167214540723

Carney, D. R., Cuddy, A. J. C., and Yap, A. J. (2010). Power posing: brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychol. Sci.* 21, 1363–1368. doi: 10.1177/0956797610383437

Casadevall, A., and Fang, F. C. (2012). Reforming science: methodological and cultural reforms. *Infect. Immunity* 80, 891–896. doi: 10.1128/IAI.06183-11

Clark, H. H. (1973). The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *J. Verb. Learn. Verb. Behav.* 12, 335–359. doi: 10.1016/S0022-5371(73)80014-3

De Houwer, J., Gawronski, B., and Barnes-Holmes, D. (2013). A functional-cognitive framework for attitude research. *Eur. Rev. Soc. Psychol.* 24, 252–287. doi: 10.1080/10463283.2014.892320

Dettweiler, U. (2019). The rationality of science and the inevitability of defining prior beliefs in empirical research. *Front. Psychol.* 10:1866. doi: 10.3389/fpsyg.2019.01866

Deutsche Forschungsgemeinschaft, (2020). *Guidelines for Publication Lists*. Available online at: https://www.dfg.de/formulare/1_91/index.jsp (accessed December 30, 2020).

Eagly, A. H., and Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychol. Rev.* 109, 573–598. doi: 10.1037/0033-295x.109.3.573

Fanelli, D. (2010). Positive" results increase down the hierarchy of the sciences. *PLoS One* 5:e10068. doi: 10.1371/journal.pone.0010068

Feldon, D., Peugh, J., Maher, M., Roksa, J., and Tofel-Grehl, C. (2017). Time-to-credit gender inequities of first-year PhD students in the biological sciences. *CBE Life Sci. Educ.* 16:a4. doi: 10.1187/cbe.16-08-0237

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychol. Rev.* 107, 659–676. doi: 10.1037/0033-295X.107.4.659

Fiedler, K. (2017). What constitutes strong psychological science? The (Neglected) role of diagnosticity and a priori theorizing. *Perspect. Psychol. Sci.* 12, 46–61. doi: 10.1177/1745691616654458

Fleck, L. (1980). *Entstehung und Entwicklung Einer Wissenschaftlichen Tatsache: Einführung in die Lehre vom Denkstil und Denkkollektiv*, 3rd Edn. Berlin: Suhrkamp Verlag.

Fox, C. W., and Paine, C. E. T. (2019). Gender differences in peer review outcomes and manuscript impact at six journals of ecology and evolution. *Ecol. Evol.* 9, 3599–3619. doi: 10.1002/ece3.4993

Giner-Sorolla, R. (2012). Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspect. Psychol. Sci.* 7, 562–571. doi: 10.1177/1745691612457576

Gino, F., and Ariely, D. (2012). The dark side of creativity: original thinkers can be more dishonest. *J. Pers. Soc. Psychol.* 102, 445–459. doi: 10.1037/a0026406

Gino, F., and Wiltermuth, S. S. (2014). Evil genius? How dishonesty can lead to greater creativity. *Psychol. Sci.* 25, 973–981. doi: 10.1177/0956797614520714

Glöckner, A., and Betsch, T. (2011). The Empirical content of theories in judgment and decision making: shortcomings and remedies. *Judgment Decis. Making* 6, 711–721.

Hagger, M. S. (2014). Avoiding the "déjà-variable" phenomenon: social psychology needs more guides to constructs. *Front. Psychol.* 5:52. doi: 10.3389/fpsyg.2014.00052

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Res. Organ. Behav.* 32, 113–135. doi: 10.1016/j.riob.2012.11.003

Hengel, E. (2017). *Publishing While Female. Are Women Held to Higher Standards? Evidence From Peer Review*. [Working Paper]. Cambridge: University of Cambridge. doi: 10.17863/CAM.17548

Henrich, J., Heine, S. J., and Norenzayan, A. (2010). The weirdest people in the world? *Behav. Brain Sci.* 33, 61–83. doi: 10.1017/S0140525X0999152X

Holzkamp, K. (1981). *Theorie und Experiment in der Psychologie: Eine grundlagenkritische Untersuchung*, 2nd Edn. Berlin: De Gruyter Studienbuch. doi: 10.1017/s0140525x0999152x

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Med.* 2:e124. doi: 10.1371/journal.pmed.0020124

Jonas, K. J., Cesario, J., Alger, M., Bailey, A. H., Bombari, D., Carney, D., et al. (2017). Power poses – where do we stand? *Compr. Results Soc. Psychol.* 2, 139–141. doi: 10.1080/23743603.2017.1342447

Judd, C. M., Westfall, J., and Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: a new and comprehensive solution to a pervasive but largely ignored problem. *J. Pers. Soc. Psychol.* 103, 54–69. doi: 10.1037/a0028347

Kazén, M., and Quirin, M. (2018). "The integration of motivation and volition in personality systems interactions (PSI) theory," in *Why People Do the Things They Do: Building on Julius Kuhl's Contributions to the Psychology of Motivation and Volition*, eds N. Baumann, M. Kazén, M. Quirin, and S. L. Koole, (Göttingen: Hogrefe), 15–30.

Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Pers. Soc. Psychol. Rev.* 2, 196–217. doi: 10.1207/s15327957pspr0203_4

Khessina, O. M., Goncalo, J. A., and Krause, V. (2018). It's time to sober up: the direct costs, side effects and long-term consequences of creativity and innovation. *Res. Organ. Behav.* 38, 107–135. doi: 10.1016/j.riob.2018.11.003

Kruschke, J. (2014). *Doing Bayesian Data Analysis: A Tutorial with R, JAGS, and Stan*, 2nd Edn. Cambridge, MA: Academic Press.

Kuhn, T. S. (1970). "Reflections on my critics," in *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science*, eds I. Lakatos, and A. Musgrave, (Cambridge: Cambridge University Press), 244–245.

Kuhn, T. S. (1996). *The Structure of Scientific Revolutions*, 3rd Edn. Chicago, IL: University of Chicago Press.

Kumar, M. (2009). A review of the review process: manuscript peer-review in biomedical research. *Biol. Med.* 1, 1–16.

Lakatos, I. (1970). "Falsification and the methodology of scientific research programm es," in *Criticism and the Growth of Knowledge*, eds I. Lakatos, and A. Musgrave, (Cambridge: Cambridge University Press).

Landers, R. N., and Behrend, T. S. (2015). an inconvenient truth: arbitrary distinctions between organizational, mechanical turk, and other convenience samples. *Indust. Organ. Psychol.* 8, 142–164. doi: 10.1017/iop.2015.13

Latour, B., and Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts*, ed. J. Salk, (Princeton, NJ: Princeton University Press).

Lerchenmueller, M. J., Sorenson, O., and Jena, A. B. (2019). Gender differences in how scientists present the importance of their research: observational study. *BMJ* 367, 60–S62. doi: 10.1136/bmj.l6573

Maxwell, S. E., Lau, M. Y., and Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does "failure to replicate" really mean? *Am. Psychol.* 70, 487–498. doi: 10.1037/a0039400

Merton, R. K. (1979). in *The Sociology of Science: Theoretical and Empirical Investigations*, ed. N. W. Storer, (Chicago, IL: University of Chicago Press).

Noah, T., Schul, Y., and Mayo, R. (2018). When both the original study and its failed replication are correct: feeling observed eliminates the facial-feedback effect. *J. Pers. Soc. Psychol.* 114, 657–664. doi: 10.1037/pspa0000121

Norton, M. I., Vandello, J. A., and Darley, J. M. (2004). Casuistry and social category bias. *J. Pers. Soc. Psychol.* 87, 817–831. doi: 10.1037/0022-3514.87.6.817

Nosek, B. A., and Bar-Anan, Y. (2012). Scientific Utopia: I. Opening scientific communication. *Psychol. Inquiry* 23, 217–243. doi: 10.1080/1047840X.2012.692215

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., and Mellor, D. T. (2018). The preregistration revolution. *Proc. Natl. Acad. Sci. U.S.A.* 115:2600. doi: 10.1073/pnas.1708274114

Nosek, B. A., Spies, J. R., and Motyl, M. (2012). Scientific Utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspect. Psychol. Sci.* 7, 615–631. doi: 10.1177/1745691612459058

Nuzzo, R. (2015). How scientists fool themselves – and how they can stop. *Nat. News* 526:182. doi: 10.1038/526182a

Open Science Collaboration, (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716

Organization Science, (2020). *Submission Guidelines | Organization Science*. Available online at: https://pubsonline.informs.org/page/orsc/submission-guidelines (accessed January 17, 2020).

Pickett, C. (2017). Let's look at the big picture: a system-level approach to assessing scholarly merit. *PsyArXiv*. Available online at: https://psyarxiv.com/tv6nb/https://psyarxiv.com/tv6nb/ (accessed January 14, 2021).

Popper, K. R. (2002). *Logik der Forschung*. Heidelberg: Mohr Siebeck.

Ray, D. G., Gomillion, S., Pintea, A. I., and Hamlin, I. (2019). On being forgotten: memory and forgetting serve as signals of interpersonal importance. *J. Pers. Soc. Psychol.* 116, 259–276. doi: 10.1037/pspi0000145

Roberson, Q. M., Phillips, K. W., Duguid, M., Thomas-Hunt, M., and Uparna, J. (2013). *Diversity as Knowledge Exchange: The Roles of Information Processing, Expertise, and Status*. Oxford: Oxford University Press.

Rouder, J. N. (2015). *Jeff Rouder // Distinguished Lecture Series in Psychology*. Available online at: https://events.psych.missouri.edu/speaker/jeff-rouder (October 15, 2015).

Rouder, J. N., and Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychon. Bull. Rev.* 18, 682–689. doi: 10.3758/s13423-011-0088-7

Salin, D. (2003). Ways of explaining workplace bullying: a review of enabling, motivating and precipitating structures and processes in the work environment. *Human Relations* 56, 1213–1232. doi: 10.1177/00187267035610003

Sherman, J. W., and Bessenoff, G. R. (1999). Stereotypes as source-monitoring cues: on the interaction between episodic and semantic memory. *Psychol. Sci.* 10, 106–110. doi: 10.1111/1467-9280.00116

Sherman, J. W., Groom, C., Mierke, K., and Klauer, K. (2003). Bearing false witness under pressure: implicit and explicit components of stereotype-driven memory distortions. *Soc. Cogn.* 21, 213–246. doi: 10.1521/soco.21.3.213.25340

Sherman, J. W., and Klein, S. A. W. (in press). The four original sins of implicit attitude research. *Front. Psychol.*

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* 22, 1359–1366. doi: 10.1177/0956797611417632

Smith, P., and Hofmann, W. (2016). Power in everyday life. *Proc. Natl. Acad. Sci. U.S.A.* 113, 10043–10048. doi: 10.1073/pnas.1604820113

Stanley, D. J., and Spence, J. R. (2014). Expectations for replications: are yours realistic? *Perspect. Psychol. Sci.* 9, 305–318. doi: 10.1177/1745691614528518

Stroebe, W., Postmes, T., and Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspect. Psychol. Sci.* 7, 670–688. doi: 10.1177/1745691612460687

Uhlmann, E. L., and Cohen, G. L. (2005). Constructed criteria: redefining merit to justify discrimination. *Psychol. Sci.* 16, 474–480. doi: 10.1111/j.0956-7976.2005.01559.x

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., and van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: the case of psi: comment on Bem (2011). *J. Pers. Soc. Psychol.* 100, 426–432. doi: 10.1037/a0022790

Wicherts, J., Kievit, R., Bakker, M., and Borsboom, D. (2012). Letting the daylight in: reviewing the reviewers and other ways to maximize transparency in science. *Front. Comp. Neurosci.* 6:20. doi: 10.3389/fncom.2012.00020

Zwebner, Y., Sellier, A.-L., Rosenfeld, N., Goldenberg, J., and Mayo, R. (2017). We look like our names: the manifestation of name stereotypes in facial appearance. *J. Pers. Soc. Psychol.* 112, 527–554. doi: 10.1037/pspa0000076.supp

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

CONCEPTUAL ANALYSIS
published: 05 February 2021
doi: 10.3389/fpsyg.2021.605191

Quantitative and Qualitative Approaches to Generalization and Replication–A Representationalist View

Matthias Borgstede and Marcel Scholz*

Foundations of Education, University of Bamberg, Bamberg, Germany

In this paper, we provide a re-interpretation of qualitative and quantitative modeling from a representationalist perspective. In this view, both approaches attempt to construct abstract representations of empirical relational structures. Whereas quantitative research uses variable-based models that abstract from individual cases, qualitative research favors case-based models that abstract from individual characteristics. Variable-based models are usually stated in the form of quantified sentences (scientific laws). This syntactic structure implies that sentences about individual cases are derived using deductive reasoning. In contrast, case-based models are usually stated using context-dependent existential sentences (qualitative statements). This syntactic structure implies that sentences about other cases are justifiable by inductive reasoning. We apply this representationalist perspective to the problems of generalization and replication. Using the analytical framework of modal logic, we argue that the modes of reasoning are often not only applied to the context that has been studied empirically, but also on a between-contexts level. Consequently, quantitative researchers mostly adhere to a top-down strategy of generalization, whereas qualitative researchers usually follow a bottom-up strategy of generalization. Depending on which strategy is employed, the role of replication attempts is very different. In deductive reasoning, replication attempts serve as empirical tests of the underlying theory. Therefore, failed replications imply a faulty theory. From an inductive perspective, however, replication attempts serve to explore the scope of the theory. Consequently, failed replications do not question the theory *per se*, but help to shape its boundary conditions. We conclude that quantitative research may benefit from a bottom-up generalization strategy as it is employed in most qualitative research programs. Inductive reasoning forces us to think about the boundary conditions of our theories and provides a framework for generalization beyond statistical testing. In this perspective, failed replications are just as informative as successful replications, because they help to explore the scope of our theories.

Keywords: qualitative research, representational measurement, research methodology, modal logic, generalizability, replication crisis

OPEN ACCESS

Edited by:
Barbara Hanfstingl,
University of Klagenfurt, Austria

Reviewed by:
Fernando Marmolejo-Ramos,
University of South Australia, Australia
Sarah Foley,
University College Cork, Ireland

****Correspondence:***
Matthias Borgstede
matthias.borgstede@uni-bamberg.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *11 September 2020*
Accepted: *11 January 2021*
Published: *05 February 2021*

Citation:
Borgstede M and Scholz M (2021) Quantitative and Qualitative Approaches to Generalization and Replication–A Representationalist View. Front. Psychol. 12:605191. doi: 10.3389/fpsyg.2021.605191

INTRODUCTION

Qualitative and quantitative research strategies have long been treated as opposing paradigms. In recent years, there have been attempts to integrate both strategies. These "mixed methods" approaches treat qualitative and quantitative methodologies as complementary, rather than opposing, strategies (Creswell, 2015). However, whilst acknowledging that both strategies have their benefits, this "integration" remains purely pragmatic. Hence, mixed methods methodology does not provide a conceptual unification of the two approaches.

Lacking a common methodological background, qualitative and quantitative research methodologies have developed rather distinct standards with regard to the aims and scope of empirical science (Freeman et al., 2007). These different standards affect the way researchers handle contradictory empirical findings. For example, many empirical findings in psychology have failed to replicate in recent years (Klein et al., 2014; Open Science, Collaboration, 2015). This "replication crisis" has been discussed on statistical, theoretical and social grounds and continues to have a wide impact on quantitative research practices like, for example, open science initiatives, pre-registered studies and a re-evaluation of statistical significance testing (Everett and Earp, 2015; Maxwell et al., 2015; Shrout and Rodgers, 2018; Trafimow, 2018; Wiggins and Chrisopherson, 2019).

However, qualitative research seems to be hardly affected by this discussion. In this paper, we argue that the latter is a direct consequence of how the concept of generalizability is conceived in the two approaches. Whereas most of quantitative psychology is committed to a top-down strategy of generalization based on the idea of random sampling from an abstract population, qualitative studies usually rely on a bottom-up strategy of generalization that is grounded in the successive exploration of the field by means of theoretically sampled cases.

Here, we show that a common methodological framework for qualitative and quantitative research methodologies is possible. We accomplish this by introducing a formal description of quantitative and qualitative models from a representationalist perspective: both approaches can be reconstructed as special kinds of representations for empirical relational structures. We then use this framework to analyze the generalization strategies used in the two approaches. These turn out to be logically independent of the type of model. This has wide implications for psychological research. First, a top-down generalization strategy is compatible with a qualitative modeling approach. This implies that mainstream psychology may benefit from qualitative methods when a numerical representation turns out to be difficult or impossible, without the need to commit to a "qualitative" philosophy of science. Second, quantitative research may exploit the bottom-up generalization strategy that is inherent to many qualitative approaches. This offers a new perspective on unsuccessful replications by treating them not as scientific failures, but as a valuable source of information about the scope of a theory.

THE QUANTITATIVE STRATEGY–NUMBERS AND FUNCTIONS

Quantitative science is about finding valid mathematical representations for empirical phenomena. In most cases, these mathematical representations have the form of functional relations between a set of variables. One major challenge of quantitative modeling consists in constructing valid measures for these variables. Formally, to measure a variable means to construct a numerical representation of the underlying empirical relational structure (Krantz et al., 1971). For example, take the behaviors of a group of students in a classroom: "to listen," "to take notes," and "to ask critical questions." One may now ask whether is possible to assign numbers to the students, such that the relations between the assigned numbers are of the same kind as the relations between the values of an underlying variable, like e.g., "engagement." The observed behaviors in the classroom constitute an empirical relational structure, in the sense that for every student-behavior tuple, one can observe whether it is true or not. These observations can be represented in a person × behavior matrix[1] (compare **Figure 1**). Given this relational structure satisfies certain conditions (i.e., the *axioms* of a measurement model), one can assign numbers to the students and the behaviors, such that the relations between the numbers resemble the corresponding numerical relations. For example, if there is a unique ordering in the empirical observations with regard to which person shows which behavior, the assigned numbers have to constitute a corresponding unique ordering, as well. Such an ordering coincides with the person × behavior matrix forming a triangle shaped relation and is formally represented by a Guttman scale (Guttman, 1944). There are various measurement models available for different empirical structures (Suppes et al., 1971). In the case of probabilistic relations, Item-Response models may be considered as a special kind of measurement model (Borsboom, 2005).

Although essential, measurement is only the first step of quantitative modeling. Consider a slightly richer empirical structure, where we observe three additional behaviors: "to doodle," "to chat," and "to play." Like above, one may ask, whether there is a unique ordering of the students with regard to these behaviors that can be represented by an underlying variable (i.e., whether the matrix forms a Guttman scale). If this is the case, we may assign corresponding numbers to the students and call this variable "distraction." In our example, such a representation is possible. We can thus assign two numbers to each student, one representing his or her "engagement" and one representing his or her "distraction" (compare **Figure 2**). These measurements can now be used to construct a quantitative model by relating the two variables by a mathematical function. In the simplest case, this may be a linear function. This functional relation constitutes a quantitative model of the empirical relational structure under

[1]A person × behavior matrix constitutes a very simple relational structure that is common in psychological research. This is why it is chosen here as a minimal example. However, more complex structures are possible, e.g., by relating individuals to behaviors over time, with individuals nested within groups etc. For a systematic overview, compare Coombs (1964).

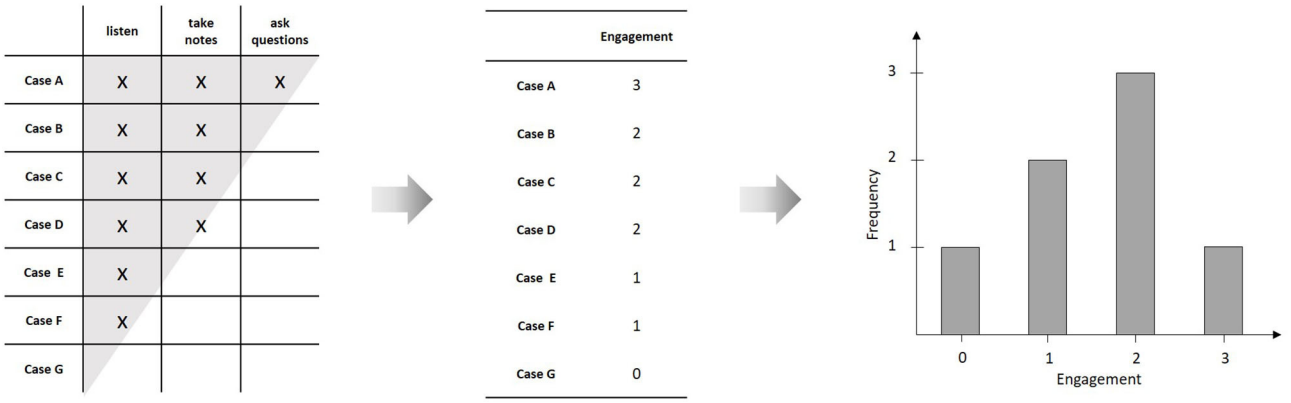


FIGURE 1 | Constructing a numerical representation from an empirical relational structure; Due to the unique ordering of persons with regard to behaviors (indicated by the triangular shape of the relation), it is possible to construct a Guttman scale by assigning a number to each of the individuals, representing the number of relevant behaviors shown by the individual. The resulting variable ("engagement") can then be described by means of statistical analyses, like, e.g., plotting the frequency distribution.

study (like, e.g., linear regression). Given the model equation and the rules for assigning the numbers (i.e., the instrumentations of the two variables), the set of admissible empirical structures is limited from all possible structures to a rather small subset. This constitutes the empirical content of the model[2] (Popper, 1935).

THE QUALITATIVE STRATEGY–CATEGORIES AND TYPOLOGIES

The predominant type of analysis in qualitative research consists in category formation. By constructing descriptive systems for empirical phenomena, it is possible to analyze the underlying empirical structure at a higher level of abstraction. The resulting categories (or types) constitute a conceptual frame for the interpretation of the observations. Qualitative researchers differ considerably in the way they collect and analyze data (Miles et al., 2014). However, despite the diverse research strategies followed by different qualitative methodologies, from a formal perspective, most approaches build on some kind of categorization of cases that share some common features. The process of category formation is essential in many qualitative methodologies, like, for example, qualitative content analysis, thematic analysis, grounded theory (see Flick, 2014 for an overview). Sometimes these features are directly observable (like in our classroom example), sometimes they are themselves the result of an interpretative process (e.g., Scheunpflug et al., 2016).

In contrast to quantitative methodologies, there have been little attempts to formalize qualitative research strategies (compare, however, Rihoux and Ragin, 2009). However, there are several statistical approaches to non-numerical data that deal with constructing abstract categories and establishing relations between these categories (Agresti, 2013). Some of these methods are very similar to qualitative category formation on a conceptual level. For example, cluster analysis groups cases into homogenous categories (clusters) based on their similarity on a distance metric.

Although category formation can be formalized in a mathematically rigorous way (Ganter and Wille, 1999), qualitative research hardly acknowledges these approaches.[3] However, in order to find a common ground with quantitative science, it is certainly helpful to provide a formal interpretation of category systems.

Let us reconsider the above example of students in a classroom. The quantitative strategy was to assign numbers to the students with regard to variables and to relate these variables via a mathematical function. We can analyze the same empirical structure by grouping the behaviors to form abstract categories. If the aim is to construct an empirically valid category system, this grouping is subject to constraints, analogous to those used to specify a measurement model. The first and most important constraint is that the behaviors must form equivalence classes, i.e., within categories, behaviors need to be equivalent, and across categories, they need to be distinct (formally, the relational structure must obey the axioms of an equivalence relation). When objects are grouped into equivalence classes, it is essential to specify the criterion for empirical equivalence. In qualitative methodology, this is sometimes referred to as the *tertium comparationis* (Flick, 2014). One possible criterion is to group behaviors such that they constitute a set of specific common attributes of a group of people. In our example, we might group the behaviors "to listen," "to take notes," and "to doodle," because these behaviors are common to the cases B, C, and D, and they are also specific for these cases, because no other person shows this particular combination of behaviors. The

[2]This notion of empirical content applies only to deterministic models. The empirical content of a probabilistic model consists in the probability distribution over all possible empirical structures.

[3]For example, neither the SAGE Handbook of qualitative data analysis edited by Flick (2014) nor the Oxford Handbook of Qualitative Research edited by Leavy (2014) mention formal approaches to category formation.

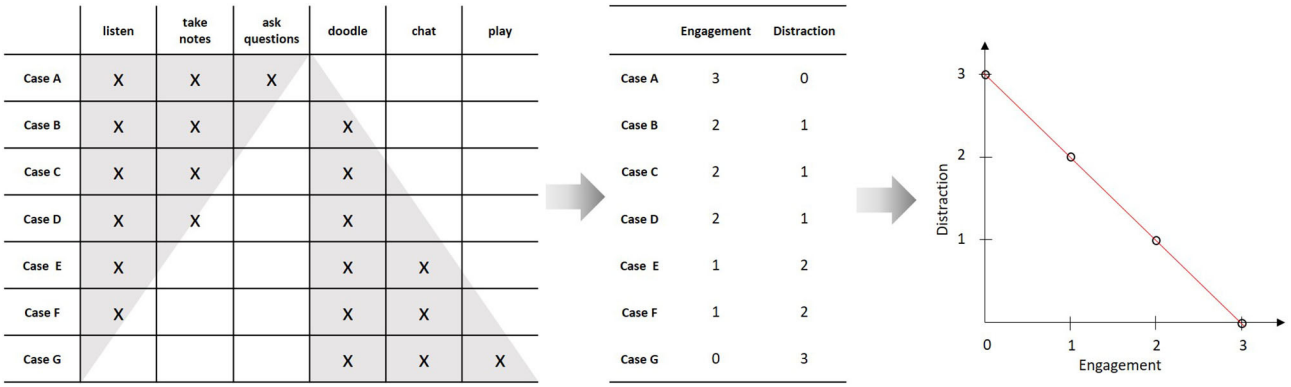


FIGURE 2 | Constructing a numerical model from an empirical relational structure; Since there are two distinct classes of behaviors that each form a Guttman scale, it is possible to assign two numbers to each individual, correspondingly. The resulting variables ("engagement" and "distraction") can then be related by a mathematical function, which is indicated by the scatterplot and red line on the right hand side.

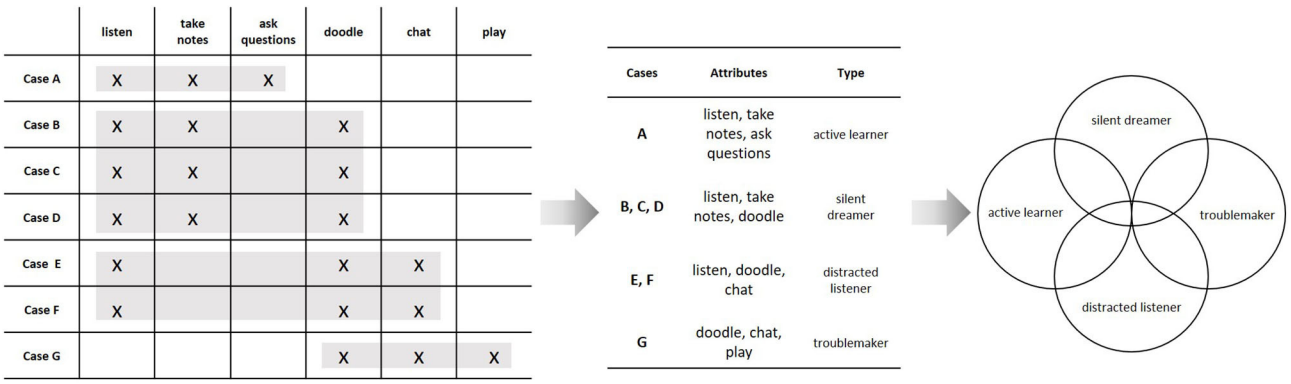


FIGURE 3 | Constructing a conceptual model from an empirical relational structure; Individual behaviors are grouped to form abstract *types* based on them being shared among a specific subset of the cases. Each type constitutes a set of specific commonalities of a class of individuals (this is indicated by the rectangles on the left hand side). The resulting types ("active learner," "silent dreamer," "distracted listener," and "troublemaker") can then be related to one another to explicate their semantic and empirical overlap, as indicated by the Venn-diagram on the right hand side.

set of common behaviors then forms an abstract concept (e.g., "moderate distraction"), while the set of persons that show this configuration form a type (e.g., "the silent dreamer"). Formally, this means to identify the *maximal rectangles* in the underlying empirical relational structure (see **Figure 3**). This procedure is very similar to the way we constructed a Guttman scale, the only difference being that we now use different aspects of the empirical relational structure.[4] In fact, the set of maximal rectangles can be determined by an automated algorithm (Ganter, 2010), just like the dimensionality of an empirical structure can be explored by psychometric scaling methods. Consequently, we can identify the empirical content of a category system or a typology as the set of empirical structures that conforms to it.[5] Whereas the quantitative strategy was to search for scalable sub-matrices and then relate the constructed variables by a mathematical function, the qualitative strategy is to construct an empirical typology by grouping cases based on their specific similarities. These types can then be related to one another by a conceptual model that describes their semantic and empirical overlap (see **Figure 3**, right hand side).

VARIABLE-BASED MODELS AND CASE-BASED MODELS

In the previous section, we have argued that qualitative category formation and quantitative measurement can both be characterized as methods to construct abstract representations of empirical relational structures. Instead of focusing on different

[4]Note also that the described structure is empirically richer than a nominal scale. Therefore, a reduction of qualitative category formation to be a special (and somehow trivial) kind of measurement is not adequate.

[5]It is possible to extend this notion of empirical content to the probabilistic case (this would correspond to applying a latent class analysis). But, since qualitative research usually does not rely on formal algorithms (neither deterministic nor probabilistic), there is currently little practical use of such a concept.

TABLE 1 | Variable-based models and case-based models.

Variable-based models	Case-based models
Primarily used in quantitative research	Primarily used in qualitative research
Description of behaviors based on person-space	Description of individuals based on behavior-space
Abstraction from individuals to populations	Abstraction from behaviors to categories
Syntactic form: $\forall_i : y_i = f(x_i)$	Syntactic form: $\exists_i : XYZ_i$
Application to behaviors: induction	Application to behaviors: deduction
Application to cases: deduction	Application to cases: induction

philosophical approaches to empirical science, we tried to stress the formal similarities between both approaches. However, it is worth also exploring the dissimilarities from a formal perspective.

Following the above analysis, the quantitative approach can be characterized by the use of variable-based models, whereas the qualitative approach is characterized by case-based models (Ragin, 1987). Formally, we can identify the rows of an empirical person × behavior matrix with a person-space, and the columns with a corresponding behavior-space. A variable-based model abstracts from the single individuals in a person-space to describe the structure of behaviors on a population level. A case-based model, on the contrary, abstracts from the single behaviors in a behavior-space to describe individual case configurations on the level of abstract categories (see **Table 1**).

From a representational perspective, there is no a priori reason to favor one type of model over the other. Both approaches provide different analytical tools to construct an abstract representation of an empirical relational structure. However, since the two modeling approaches make use of different information (person-space vs. behavior-space), this comes with some important implications for the researcher employing one of the two strategies. These are concerned with the role of deductive and inductive reasoning.

In variable-based models, empirical structures are represented by functional relations between variables. These are usually stated as scientific laws (Carnap, 1928). Formally, these laws correspond to logical expressions of the form

$$\forall_i : y_i = f(x_i)$$

In plain text, this means that y is a function of x for all objects i in the relational structure under consideration. For example, in the above example, one may formulate the following law: for all students in the classroom it holds that "distraction" is a monotone decreasing function of "engagement." Such a law can be used to derive predictions for single individuals by means of logical deduction: if the above law applies to all students in the classroom, it is possible to calculate the expected distraction from a student's engagement. An empirical observation can now be evaluated against this prediction. If the prediction turns out to be false, the law can be refuted based on the principle of falsification (Popper, 1935). If a scientific law repeatedly withstands such empirical tests, it may be considered to be valid with regard to the relational structure under consideration.

In case-based models, there are no laws about a population, because the model does not abstract from the cases but from the observed behaviors. A case-based model describes the underlying structure in terms of existential sentences. Formally, this corresponds to a logical expression of the form

$$\exists_i : XYZ_i$$

In plain text, this means that there is at least one case i for which the condition XYZ holds. For example, the above category system implies that there is at least one active learner. This is a statement about a singular observation. It is impossible to deduce a statement about another person from an existential sentence like this. Therefore, the strategy of falsification cannot be applied to test the model's validity in a specific context. If one wishes to generalize to other cases, this is accomplished by inductive reasoning, instead. If we observed one person that fulfills the criteria of calling him or her an active learner, we can hypothesize that there may be other persons that are identical to the observed case in this respect. However, we do not arrive at this conclusion by logical deduction, but by induction.

Despite this important distinction, it would be wrong to conclude that variable-based models are intrinsically deductive and case-based models are intrinsically inductive.[6] Both types of reasoning apply to both types of models, but on different levels. Based on a person-space, in a variable-based model one can use deduction to derive statements about individual persons from abstract population laws. There is an analogous way of reasoning for case-based models: because they are based on a behavior space, it is possible to deduce statements about singular behaviors. For example, if we know that Peter is an active learner, we can deduce that he takes notes in the classroom. This kind of deductive reasoning can also be applied on a higher level of abstraction to deduce thematic categories from theoretical assumptions (Braun and Clarke, 2006). Similarly, there is an analog for inductive generalization from the perspective of variable-based modeling: since the laws are only quantified over the person-space, generalizations to other behaviors rely on inductive reasoning. For example, it is plausible to assume that highly engaged students tend to do their homework properly–however, in our example this behavior has never been observed. Hence, in variable-based models we usually generalize to other behaviors by means of induction. This kind of inductive reasoning is very common when empirical results are generalized from the laboratory to other behavioral domains.

Although inductive and deductive reasoning are used in qualitative and quantitative research, it is important to stress the different roles of induction and deduction when models are applied to cases. A variable-based approach implies to draw

[6]We do not elaborate on abductive reasoning here, since, given an empirical relational structure, the concept can be applied to both types of models in the same way (Schurz, 2008). One could argue that the underlying relational structure is not given a priori but has to be constructed by the researcher and will itself be influenced by theoretical expectations. Therefore, abductive reasoning may be necessary to establish an empirical relational structure in the first place.

conclusions about cases by means of logical deduction; a case-based approach implies to draw conclusions about cases by means of inductive reasoning. In the following, we build on this distinction to differentiate between qualitative (bottom-up) and quantitative (top-down) strategies of generalization.

GENERALIZATION AND THE PROBLEM OF REPLICATION

We will now extend the formal analysis of quantitative and qualitative approaches to the question of generalization and replicability of empirical findings. For this sake, we have to introduce some concepts of formal logic. Formal logic is concerned with the validity of arguments. It provides conditions to evaluate whether certain sentences (conclusions) can be derived from other sentences (premises). In this context, a *theory* is nothing but a set of sentences (also called axioms). Formal logic provides tools to derive new sentences that must be true, given the axioms are true (Smith, 2020). These derived sentences are called *theorems* or, in the context of empirical science, *predictions* or *hypotheses*. On the syntactic level, the rules of logic only state how to evaluate the truth of a sentence relative to its premises. Whether or not sentences are actually true, is formally specified by logical semantics.

On the semantic level, formal logic is intrinsically linked to set-theory. For example, a logical statement like "all dogs are mammals," is true if and only if the set of dogs is a subset of the set of mammals. Similarly, the sentence "all chatting students doodle" is true if and only if the set of chatting students is a subset of the set of doodling students (compare **Figure 3**). Whereas, the first sentence is analytically true due to the way we define the words "dog" and "mammal," the latter can be either true or false, depending on the relational structure we actually observe. We can thus interpret an empirical relational structure as the truth criterion of a scientific theory. From a logical point of view, this corresponds to the semantics of a theory. As shown above, variable-based and case-based models both give a formal representation of the same kinds of empirical structures. Accordingly, both types of models can be stated as formal theories. In the variable-based approach, this corresponds to a set of scientific laws that are quantified over the members of an abstract population (these are the axioms of the theory). In the case-based approach, this corresponds to a set of abstract existential statements about a specific class of individuals.

In contrast to mathematical axiom systems, empirical theories are usually not considered to be *necessarily* true. This means that even if we find no evidence against a theory, it is still possible that it is actually wrong. We may know that a theory is valid in some contexts, yet it may fail when applied to a new set of behaviors (e.g., if we use a different instrumentation to measure a variable) or a new population (e.g., if we draw a new sample).

From a logical perspective, the possibility that a theory may turn out to be false stems from the problem of *contingency*. A statement is contingent, if it is both, possibly true and possibly false. Formally, we introduce two *modal* operators: $\Box$ to designate logical necessity, and $\Diamond$ to designate logical possibility. Semantically, these operators are very similar to the existential quantifier, $\exists$, and the universal quantifier, $\forall$. Whereas $\exists$ and $\forall$ refer to the individual objects within one relational structure, the modal operators $\Box$ and $\Diamond$ range over so-called *possible worlds*: a statement is possibly true, if and only if it is true in at least one accessible possible world, and a statement is necessarily true if and only if it is true in every accessible possible world (Hughes and Cresswell, 1996). Logically, possible worlds are mathematical abstractions, each consisting of a relational structure. Taken together, the relational structures of all accessible possible worlds constitute the formal semantics of necessity, possibility and contingency.[7]

In the context of an empirical theory, each possible world may be identified with an empirical relational structure like the above classroom example. Given the set of intended applications of a theory (the *scope* of the theory, one may say), we can now construct possible world semantics for an empirical theory: each intended application of the theory corresponds to a possible world. For example, a quantified sentence like "all chatting students doodle" may be true in one classroom and false in another one. In terms of possible worlds, this would correspond to a statement of contingency: "it is possible that all chatting students doodle in one classroom, and it is possible that they don't in another classroom." Note that in the above expression, "all students" refers to the students in only one possible world, whereas "it is possible" refers to the fact that there is at least one possible world for each of the specified cases.

To apply these possible world semantics to quantitative research, let us reconsider how generalization to other cases works in variable-based models. Due to the syntactic structure of quantitative laws, we can deduce predictions for singular observations from an expression of the form $\forall_i : y_i = f(x_i)$. Formally, the logical quantifier $\forall$ ranges only over the objects of the corresponding empirical relational structure (in our example this would refer to the students in the observed classroom). But what if we want to generalize beyond the empirical structure we actually observed? The standard procedure is to assume an infinitely large, abstract population from which a random sample is drawn. Given the truth of the theory, we can deduce predictions about what we may observe in the sample. Since usually we deal with probabilistic models, we can evaluate our theory by means of the conditional probability of the observations, given the theory holds. This concept of conditional probability is the foundation of statistical significance tests (Hogg et al., 2013), as well as Bayesian estimation (Watanabe, 2018). In terms of possible world semantics, the random sampling model implies that all possible worlds (i.e., all intended applications) can be conceived as empirical sub-structures from a greater population structure. For example, the empirical relational structure constituted by the observed behaviors in a classroom would be conceived as a sub-matrix of the population

[7]We shall not elaborate on the metaphysical meaning of possible worlds here, since we are only concerned with empirical theories [but see Tooley (1999), for an overview].

person × behavior matrix. It follows that, if a scientific law is true in the population, it will be true in all possible worlds, i.e., it will be *necessarily* true. Formally, this corresponds to an expression of the form

$$\Box(\forall_i : y_i = f(x_i))$$

The statistical generalization model thus constitutes a top-down strategy for dealing with individual contexts that is analogous to the way variable-based models are applied to individual cases (compare **Table 1**). Consequently, if we apply a variable-based model to a new context and find out that it does not fit the data (i.e., there is a statistically significant deviation from the model predictions), we have reason to doubt the validity of the theory. This is what makes the problem of low replicability so important: we observe that the predictions are wrong in a new study; and because we apply a top-down strategy of generalization to contexts beyond the ones we observed, we see our whole theory at stake.

Qualitative research, on the contrary, follows a different strategy of generalization. Since case-based models are formulated by a set of context-specific existential sentences, there is no need for universal truth or necessity. In contrast to statistical generalization to other cases by means of random sampling from an abstract population, the usual strategy in case-based modeling is to employ a bottom-up strategy of generalization that is analogous to the way case-based models are applied to individual cases. Formally, this may be expressed by stating that the observed qualia exist in at least one possible world, i.e., the theory is *possibly* true:

$$\Diamond(\exists_i : XYZ_i)$$

This statement is analogous to the way we apply case-based models to individual cases (compare **Table 1**). Consequently, the set of intended applications of the theory does not follow from a sampling model, but from theoretical assumptions about which cases may be similar to the observed cases with respect to certain relevant characteristics. For example, if we observe that certain behaviors occur together in one classroom, following a bottom-up strategy of generalization, we will hypothesize why this might be the case. If we do not replicate this finding in another context, this does not question the model itself, since it was a context-specific theory all along. Instead, we will revise our hypothetical assumptions about why the new context is apparently less similar to the first one than we originally thought. Therefore, if an empirical finding does not replicate, we are more concerned about our understanding of the cases than about the validity of our theory.

Whereas statistical generalization provides us with a formal (and thus somehow more objective) apparatus to evaluate the universal validity of our theories, the bottom-up strategy forces us to think about the class of intended applications on theoretical grounds. This means that we have to ask: what are the boundary conditions of our theory? In the above classroom example, following a bottom-up strategy, we would build on our preliminary understanding of the cases in one context (e.g., a public school) to search for similar and contrasting cases in other contexts (e.g., a private school). We would then re-evaluate our theoretical description of the data and explore what makes cases similar or dissimilar with regard to our theory. This enables us to expand the class of intended applications alongside with the theory.

Of course, none of these strategies is superior *per se*. Nevertheless, they rely on different assumptions and may thus be more or less adequate in different contexts. The statistical strategy relies on the assumption of a universal population and invariant measurements. This means, we assume that (a) all samples are drawn from the same population and (b) all variables refer to the same behavioral classes. If these assumptions are true, statistical generalization is valid and therefore provides a valuable tool for the testing of empirical theories. The bottom-up strategy of generalization relies on the idea that contexts may be classified as being more or less similar based on characteristics that are not part of the model being evaluated. If such a similarity relation across contexts is feasible, the bottom-up strategy is valid, as well. Depending on the strategy of generalization, replication of empirical research serves two very different purposes. Following the (top-down) principle of generalization by deduction from scientific laws, replications are empirical tests of the theory itself, and failed replications question the theory on a fundamental level. Following the (bottom-up) principle of generalization by induction to similar contexts, replications are a means to explore the boundary conditions of a theory. Consequently, failed replications question the scope of the theory and help to shape the set of intended applications.

CONCLUSION

We have argued that quantitative and qualitative research are best understood by means of the structure of the employed models. Quantitative science mainly relies on variable-based models and usually employs a top-down strategy of generalization from an abstract population to individual cases. Qualitative science prefers case-based models and usually employs a bottom-up strategy of generalization. We further showed that failed replications have very different implications depending on the underlying strategy of generalization. Whereas in the top-down strategy, replications are used to test the universal validity of a model, in the bottom-up strategy, replications are used to explore the scope of a model. We will now address the implications of this analysis for psychological research with regard to the problem of replicability.

Modern day psychology almost exclusively follows a top-down strategy of generalization. Given the quantitative background of most psychological theories, this is hardly surprising. Following the general structure of variable-based

models, the individual case is not the focus of the analysis. Instead, scientific laws are stated on the level of an abstract population. Therefore, when applying the theory to a new context, a statistical sampling model seems to be the natural consequence. However, this is not the only possible strategy. From a logical point of view, there is no reason to assume that a quantitative law like $\forall_i : y_i = f(x_i)$ implies that the law is necessarily true, i.e.,: $\Box(\forall_i : y_i = f(x_i))$. Instead, one might just as well define the scope of the theory following an inductive strategy.[8] Formally, this would correspond to the assumption that the observed law is *possibly* true, i.e.,: $\Diamond(\forall_i : y_i = f(x_i))$. For example, we may discover a functional relation between "engagement" and "distraction" without referring to an abstract universal population of students. Instead, we may hypothesize under which conditions this functional relation may be valid and use these assumptions to inductively generalize to other cases.

If we take this seriously, this would require us to specify the intended applications of the theory: in which contexts do we expect the theory to hold? Or, equivalently, what are the boundary conditions of the theory? These boundary conditions may be specified either intensionally, i.e., by giving external criteria for contexts being similar enough to the ones already studied to expect a successful application of the theory. Or they may be specified extensionally, by enumerating the contexts where the theory has already been shown to be valid. These boundary conditions need not be restricted to the population we refer to, but include all kinds of contextual factors. Therefore, adopting a bottom-up strategy, we are forced to think about these factors and make them an integral part of our theories.

In fact, there is good reason to believe that bottom-up generalization may be more adequate in many psychological studies. Apart from the pitfalls associated with statistical generalization that have been extensively discussed in recent years (e.g., p-hacking, underpowered studies, publication bias), it is worth reflecting on whether the underlying assumptions are met in a particular context. For example, many samples used in experimental psychology are not randomly drawn from a large population, but are convenience samples. If we use statistical models with non-random samples, we have to assume that the observations vary *as if* drawn from a random sample. This may indeed be the case for randomized experiments, because all variation between the experimental conditions apart from the independent variable will be random due to the randomization procedure. In this case, a classical significance test may be regarded as an approximation to a randomization test (Edgington and Onghena, 2007). However, if we interpret a significance test as an approximate randomization test, we test not for generalization but for internal validity. Hence, even if we use statistical significance tests when assumptions about random sampling are violated, we still have to use a different strategy of generalization. This issue has been discussed in the context of small-N studies, where variable-based models are applied to very small samples, sometimes consisting of only one individual (Dugard et al., 2012). The bottom-up strategy of generalization that is employed by qualitative researchers, provides such an alternative.

Another important issue in this context is the question of measurement invariance. If we construct a variable-based model in one context, the variables refer to those behaviors that constitute the underlying empirical relational structure. For example, we may construct an abstract measure of "distraction" using the observed behaviors in a certain context. We will then use the term "distraction" as a theoretical term referring to the variable we have just constructed to represent the underlying empirical relational structure. Let us now imagine we apply this theory to a new context. Even if the individuals in our new context are part of the same population, we may still get into trouble if the observed behaviors differ from those used in the original study. How do we know whether these behaviors constitute the same variable? We have to ensure that in any new context, our measures are valid for the variables in our theory. Without a proper measurement model, this will be hard to achieve (Buntins et al., 2017). Again, we are faced with the necessity to think of the boundary conditions of our theories. In which contexts (i.e., for which sets of individuals and behaviors) do we expect our theory to work?

If we follow the rationale of inductive generalization, we can explore the boundary conditions of a theory with every new empirical study. We thus widen the scope of our theory by comparing successful applications in different contexts and unsuccessful applications in similar contexts. This may ultimately lead to a more general theory, maybe even one of universal scope. However, unless we have such a general theory, we might be better off, if we treat unsuccessful replications not as a sign of failure, but as a chance to learn.

AUTHOR CONTRIBUTIONS

MB conceived the original idea and wrote the first draft of the paper. MS helped to further elaborate and scrutinize the arguments. All authors contributed to the final version of the manuscript.

ACKNOWLEDGMENTS

We would like to thank Annette Scheunpflug for helpful comments on an earlier version of the manuscript.

[8]Of course, this also means that it would be equally reasonable to employ a top-down strategy of generalization using a case-based model by postulating that $\Box(\exists_i : XYZ_i)$. The implications for case-based models are certainly worth exploring, but lie beyond the scope of this article.

REFERENCES

Agresti, A. (2013). *Categorical Data Analysis, 3rd Edn. Wiley Series In Probability And Statistics*. Hoboken, NJ: Wiley.

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511490026

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Buntins, M., Buntins, K., and Eggert, F. (2017). Clarifying the concept of validity: from measurement to everyday language. *Theory Psychol.* 27, 703–710. doi: 10.1177/0959354317702256

Carnap, R. (1928). *The Logical Structure of the World*. Berkeley, CA: University of California Press.

Coombs, C. H. (1964). *A Theory of Data*. New York, NY: Wiley.

Creswell, J. W. (2015). *A Concise Introduction to Mixed Methods Research*. Los Angeles, CA: Sage.

Dugard, P., File, P., and Todman, J. B. (2012). *Single-Case and Small-N Experimental Designs: A Practical Guide to Randomization Tests 2nd Edn*. New York, NY: Routledge. doi: 10.4324/9780203180938

Edgington, E., and Onghena, P. (2007). *Randomization Tests, 4th Edn. Statistics.* Hoboken, NJ: CRC Press. doi: 10.1201/9781420011814

Everett, J. A. C., and Earp, B. D. (2015). A tragedy of the (academic) commons: interpreting the replication crisis in psychology as a social dilemma for early-career researchers. *Front. Psychol.* 6:1152. doi: 10.3389/fpsyg.2015.01152

Flick, U. (Ed.). (2014). *The Sage Handbook of Qualitative Data Analysis*. London: Sage. doi: 10.4135/9781446282243

Freeman, M., Demarrais, K., Preissle, J., Roulston, K., and St. Pierre, E. A. (2007). Standards of evidence in qualitative research: an incitement to discourse. *Educ. Res.* 36, 25–32. doi: 10.3102/0013189X06298009

Ganter, B. (2010). "Two basic algorithms in concept analysis," in *Lecture Notes In Computer Science. Formal Concept Analysis, Vol. 5986,* eds D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, F. Mattern, J. C. Mitchell, et al. (Berlin, Heidelberg: Springer Berlin Heidelberg), 312–340. doi: 10.1007/978-3-642-11928-6_22

Ganter, B., and Wille, R. (1999). *Formal Concept Analysis*. Berlin, Heidelberg: Springer Berlin Heidelberg. doi: 10.1007/978-3-642-59830-2

Guttman, L. (1944). A basis for scaling qualitative data. *Am. Sociol. Rev.* 9:139. doi: 10.2307/2086306

Hogg, R. V., Mckean, J. W., and Craig, A. T. (2013). *Introduction to Mathematical Statistics, 7th Edn*. Boston, MA: Pearson.

Hughes, G. E., and Cresswell, M. J. (1996). *A New Introduction To Modal Logic*. London; New York, NY: Routledge. doi: 10.4324/9780203290644

Klein, R. A., Ratliff, K. A., Vianello, M., Adams, R. B., Bahník, Š., Bernstein, M. J., et al. (2014). Investigating variation in replicability. *Soc. Psychol.* 45, 142–152. doi: 10.1027/1864-9335/a000178

Krantz, D. H., Luce, D., Suppes, P., and Tversky, A. (1971). *Foundations of Measurement Volume I: Additive And Polynomial Representations*. New York, NY; London: Academic Press. doi: 10.1016/B978-0-12-425401-5.50011-8

Leavy, P. (2014). *The Oxford Handbook of Qualitative Research*. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780199811755.001.0001

Maxwell, S. E., Lau, M. Y., and Howard, G. S. (2015). Is psychology suffering from a replication crisis? what does "failure to replicate" really mean? *Am. Psychol.* 70, 487–498. doi: 10.1037/a0039400

Miles, M. B., Huberman, A. M., and Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook, 3rd Edn*. Los Angeles, CA; London; New Delhi; Singapore; Washington, DC: Sage.

Open Science, Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:Aac4716. doi: 10.1126/science.aac4716

Popper, K. (1935). *Logik Der Forschung*. Wien: Springer. doi: 10.1007/978-3-7091-4177-9

Ragin, C. (1987). *The Comparative Method : Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA: University Of California Press.

Rihoux, B., and Ragin, C. (2009). *Configurational Comparative Methods: Qualitative Comparative Analysis (Qca) And Related Techniques*. Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781452226569

Scheunpflug, A., Krogull, S., and Franz, J. (2016). Understanding learning in world society: qualitative reconstructive research in global learning and learning for sustainability. *Int. Journal Dev. Educ. Glob. Learn.* 7, 6–23. doi: 10.18546/IJDEGL.07.3.02

Schurz, G. (2008). Patterns of abduction. *Synthese* 164, 201–234. doi: 10.1007/s11229-007-9223-4

Shrout, P. E., and Rodgers, J. L. (2018). Psychology, science, and knowledge construction: broadening perspectives from the replication crisis. *Annu. Rev. Psychol.* 69, 487–510. doi: 10.1146/annurev-psych-122216-011845

Smith, P. (2020). *An Introduction To Formal Logic*. Cambridge: Cambridge University Press. doi: 10.1017/9781108328999

Suppes, P., Krantz, D. H., Luce, D., and Tversky, A. (1971). *Foundations of Measurement Volume II: Geometrical, Threshold, and Probabilistic Representations*. New York, NY; London: Academic Press.

Tooley, M. (Ed.) (1999). *Necessity and Possibility. The Metaphysics of Modality*. New York, NY; London: Garland Publishing.

Trafimow, D. (2018). An a priori solution to the replication crisis. *Philos. Psychol.* 31, 1188–1214. doi: 10.1080/09515089.2018.1490707

Watanabe, S. (2018). *Mathematical Foundations of Bayesian Statistics. CRC Monographs On Statistics And Applied Probability*. Boca Raton, FL: Chapman And Hall.

Wiggins, B. J., and Chrisopherson, C. D. (2019). The replication crisis in psychology: an overview for theoretical and philosophical psychology. *J. Theor. Philos. Psychol.* 39, 202–217. doi: 10.1037/teo0000137

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

OPINION
published: 19 February 2021
doi: 10.3389/fpsyg.2021.612570

Regarding an "Almost Anything Goes" Attitude Toward Methods in Psychology

Steffen Zitzmann[1]* *and Lukas Loreth*[2]

[1] Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Tübingen, Germany, [2] Department of Psychology, Kiel University, Kiel, Germany

Keywords: methodology, statistics, disapproval, respect, tolerance

"The beauty of the universe consists not only of unity in variety,
but also of variety in unity."
– Umberto Eco

Post-modernism is a movement that recognizes and encourages different views. As such, it has shaped contemporary democratic societies. One key feature of these societies is diversity, which enriches such societies through cultural exchange and learning. However, because diversity can also lead to conflicts, there have been calls for mutual tolerance (Scanlon, 2003; Popper, 2013). In recent years, psychology has increasingly acknowledged the notion of tolerance as the attitude that one permits others to have different ways of life (i.e., their beliefs, preferences, and practices) despite one's disapproval of them (e.g., Simon, 2020; see also Verkuyten et al., 2020). Tolerance is made possible by respecting others regardless of one's disapproval of them; i.e., by recognizing them as equals (e.g., as citizens with the same rights, duties, and liberties). It is the basis for peace and also paves the way for cooperation between people across boundaries.

In this commentary, we assume that contemporary psychology can be characterized in the very same way as democratic societies. Psychology encompasses not only a variety of different subdisciplines but also proponents of different statistical approaches and methods. These researchers often disapprove of one another because they disapprove of other researchers' work, specifically the statistical approaches other researchers adhere to or the particular methods they prefer. Note that the term disapproval refers to disapproval of what these researchers stand up for. Thereby, it taps the essential, defining feature (e.g., being a Frequentist vs. being a Bayesian) and not some sort of interpersonal disliking. To exemplify such mutual disapproval among researchers, we point to two controversies that are currently heating up and that serve as examples. We then briefly discuss the impact of such controversies on research practice on the basis of our own experience as researchers in the field. Finally, as a remedy, we offer a post-modern methodology that is liberal, pluralistic, and more tolerant.

Today, two principal ways of doing statistics are currently in use in psychology. In the early days of modern statistics, the Bayesian approach—named after Thomas Bayes—took shape and played a major role in the field of statistics, whereas later in the 20th century, it was superseded by the frequentist approach, which was launched by Ronald A. Fisher and others and which many researchers have adopted in their work. Although this approach dominates in psychology, the Bayesian approach has been on the rise again in recent years (van de Schoot et al., 2017). The main difference between the two approaches, and also often the point of disagreement among researchers, is that the Bayesian approach uses not only the data at hand but also a so-called prior distribution. The prior expresses previous knowledge from a previous study, a meta-analysis, or an expert, for example. Conversely, the frequentist approach does not make use of such information, and this has led Efron (2005) to compare a researcher who adheres to the frequentist approach with "a Bayesian trying to do well, or at least not too badly, against any possible prior distribution"

OPEN ACCESS

Edited by:
Timo Gnambs,
Leibniz Institute for Educational Trajectories (LG), Germany

Reviewed by:
Hervé Guyon,
Université de Bretagne Occidentale, France

***Correspondence:**
Steffen Zitzmann
steffen.zitzmann@uni-tuebingen.de

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *30 September 2020*
Accepted: *12 January 2021*
Published: *19 February 2021*

Citation:
Zitzmann S and Loreth L (2021) Regarding an "Almost Anything Goes" Attitude Toward Methods in Psychology. Front. Psychol. 12:612570. doi: 10.3389/fpsyg.2021.612570

(p. 2). The Bayesian approach has many advantages. Among them, perhaps the two most interesting ones are the abilities to incorporate previous findings from related studies and to stabilize models by appropriately specifying the prior (Lüdtke et al., 2018; Zitzmann et al., 2020; see also Zitzmann et al., 2021). For further arguments for why the Bayesian approach might be attractive to researchers, see, e.g., Muthén and Asparouhov (2012) or Depaoli and Clifton (2015).

Besides the major controversy regarding the "right" approach to statistics in psychology, there are also minor controversies. One such controversy is related to the validity of different methods for estimating measurement models that differ in their assumptions and procedures. Specifically, researchers have held debates on which approach is most suitable: factor-based or composite-based methods. Although factor-based methods, such as common maximum likelihood factor analysis or structural equation modeling tend to dominate in psychology, proponents of composite-based methods have argued that these methods can be superior to factor-based methods. One such composite-based method is partial least squares (PLS), and scholars have emphasized the advantage of PLS when the sample size is small (Wold, 1982). This is because the method does not fit the whole model at once but first divides the model into simpler submodels, and then these submodels are fit separately (Tenenhaus et al., 2005; Zitzmann and Helm, 2021). See Rigdon et al. (2017) for an in-depth discussion. However, the fact that the Bayesian approach and PLS can be advantageous under certain conditions does not mean that the frequentist approach and the factor-based methods should be abandoned. Rather, we want to acknowledge that all methods have their undeniable strengths and all are indeed useful in practice.

Much of our motivation for writing this commentary has stemmed from our own experience in getting articles published as well as from the many discussions we have had with other researchers in the field. We believe that readers have similar experiences, although there may be nuances. Publishing an article that uses non-standard methods is still challenging because journal editors and reviewers tend to be overly critical, particularly when they favor another method. Moreover, researchers tend to overlook or even actively ignore even the published work of other researchers when this work used a different statistical approach or method. All of this hampers scientific progress, and we think researchers can and *should* do better.

Karl Popper, who is well-known by psychologists, suggested that science should be "hypothetico-deductive," meaning that researchers should scrutinize theories by deducing and then testing hypotheses that are falsifiable on empirical grounds. Not all but a great deal of research in psychology is devoted to this idea, and research articles are usually framed in this way. However, in his famous book *Against Method*, Paul Feyerabend (2010) argued that the prescription of one method could hamper science and that the spectrum of methods is much broader. He suggested that science would benefit from a mild "anarchism" (i.e., no rigid rules), which is why he was called an "anarchist."[1] We use the word "liberalism" here instead of "anarchism" because we find it more suitable for characterizing our concept of a post-modern methodology. Moreover, and more important for our proposal, Feyerabend also coined the phrase "anything goes," which we also use and with which we refer to a more pluralistic methodology. By this, we mean that adhering to a specific statistical approach or using a specific method in research practice is perfectly fine, but there are more approaches and methods out there, and researchers should tolerate the researchers who use these methods.

In recent years, the concept of tolerance has been developed in social psychology by adopting ideas from philosophy (e.g., Honneth, 1995; Forst, 2013), and this has led to the disapproval-respect model of tolerance (Simon and Schaefer, 2016; see also Simon et al., 2019; Simon, 2020). This model is a dual-level model. In accordance with Turner et al. (1987), the model assumes that social groups and their respective identities are hierarchically arranged. Members of different groups at a lower level can be members of the same superordinate group at a higher level. A common group identity at the higher level grounds a mutual recognition of equality, meaning that members of different subordinate groups can still recognize each other as equals (because they belong to the same group at the higher level). This ability has also been termed "respect for others as equals." Respecting others as equals is the main driving force for tolerance, whereas disapproval is a definitional condition for tolerance (Simon, 2020). This means that to develop mutual tolerance, people do not need to give up their disapproval (or their lower level identity). Rather, they need to be embedded in the shared superordinate identity and respect each other as equals (Simon, 2020).

This model can be applied to psychology as well. Researchers are capable of respecting other researchers as equal fellow researchers, even when they differ in their work, specifically in the statistical approaches they adhere to and the methods they use. Respect may then help researchers develop mutual tolerance (i.e., the attitude that one permits the different approaches and methods used by other researchers despite one's disapproval of them) because respect drives tolerance. It is important to note that recognizing equality does not require researchers to see similarities between different approaches or methods or require them to start liking other researchers. The task is much simpler: They need to strengthen their shared identity as psychologists. One way to further strengthen this identity is to facilitate publication in the same respected journals even for researchers whose approaches and methods are not mainstream. To this end, editors should be aware of whether their own as well as the reviewers' methodological critiques are really justified or whether they are merely an expression of dislike. To help editors,

[1]It is interesting to note that, as Hacking pointed out in the Foreword of the fourth edition of *Against Method*, Feyerabend himself preferred the term "Dadaist" for various reasons. Thus, in a letter to Imre Lakatos, Feyerabend wrote that he hoped that "the reader will remember me as a flippant Dadaist and not as a serious anarchist."

they could—if not should—select reviewers who are also familiar with non-mainstream methods. Once they are publishing in the same respected journals, researchers might become aware of the work of others and might even be influenced by such work. We would like to note that the evaluation of possible shortcomings needs some kind of standard. Whether this standard must be a common standard or whether each approach or method can only be evaluated against its own standards can be debated. We think that a common standard is not incompatible with our proposed methodology because different methods can nevertheless be evaluated and compared with each other by applying general criteria, such as statistical criteria (e.g., the accuracy of results).

Another way to strengthen researchers' shared identity is to change the way University teachers teach methods. We suggest they always be taught "in the plural" (see Gigerenzer et al., 2004), which means that non-mainstream methods should also be an element of the method curriculum. However, most University teachers are not trained in non-mainstream methods, and psychology textbooks do not show and explain these methods. As a remedy, University teachers might want to consult the literature on data science, a discipline that is intended to understand phenomena using data and a plethora of different methods to analyze them.

In sum, this commentary points out two controversies that surround methods, followed by a brief outline of a post-modern methodology in psychology. We view mutual tolerance among proponents of different statistical approaches and methods as the key feature of this methodology, which is vital for a pluralism of methods and thereby also for psychology in post-modern times because it facilitates fruitful exchange and cooperation.

We wish to emphasize that such a methodology does not necessarily imply pragmatism. Pragmatists sometimes argue that the significance of the statistical approach or method is often exaggerated because, in many applications, the practical differences between numerical results are only small, and thus, results can be interpreted in multiple ways (e.g., Albers et al., 2018). However, we think that even if results are very similar, their interpretations will nevertheless differ, and researchers may want prefer one or the other of these (different) interpretations. Therefore, we suggest that interpretations not be mixed or blurred but be in accordance with the specific approach or method that was used (see Nalborczyk et al., 2019, for a very similar argument). Moreover, our post-modern methodology does not imply that truth cannot be approached or constructed. We view the search for truth (or the construction thereof) as a guiding principle for researchers that operates as a further source of a mutual recognition of equality: Recognizing this task as a collective endeavor can help researchers give respect to others as equal fellows and tolerate them even when they disapprove of them. Also, our notion of post-modernism does not entail that any discourse is acceptable. We indicated this by the word "almost" in the phrase "almost anything goes" in the commentary's title. This means that we do not consider a radical epistemological critique. Rather, we suggest that the discourse should take place with adherence to a minimal standard (e.g., statistical criteria). First and foremost, our notion entails a call for mutual tolerance, i.e., a critique of rejecting methods because they do not conform to the dominant ideology.

To conclude, our outline points out three aspects of a new post-modern methodology in psychology: liberal, pluralistic, and more tolerant: liberal because it rejects rules that are too strict in favor of more freedom in the choice of method, pluralistic because it conveys an "almost anything goes" attitude toward methods, and more tolerant because mutual tolerance among researchers is vital for a pluralism of methods. Psychological phenomena are complex and can best be understood by using different methods (Mayrhofer and Hutmacher, 2020). However, to get things working, tolerance must actively be lived. Of course, much depends on our own willingness as researchers but also on the system's arrangements. Psychology could be more colorful, and we could all have more fun if we were to be more committed to such a methodology. We hope our commentary has offered a view on methods that both new and established researchers in psychology will find attractive.

AUTHOR CONTRIBUTIONS

SZ: writing and lead. LL: writing. Both authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We acknowledge support by Open Access Publishing Fund of University of Tübingen.

REFERENCES

Albers, C. J., Kiers, H. A. L., and van Ravenzwaaij, D. (2018). Credible confidence: a pragmatic view on the frequentist vs Bayesian debate. *Collabra Psychol*. 4, 1–8. doi: 10.1525/collabra.149

Depaoli, S., and Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Struct. Equat. Model.* 22, 327–351. doi: 10.1080/10705511.2014.937849

Efron, B. (2005). *Modern Science and the Bayesian-Frequentist Controversy*. Stanford, CA: Division of Biostatistics, Stanford University Press.

Feyerabend, P. (2010). *Against Method, 4th Edn*. New York, NY: Verso.

Forst, R. (2013). *Toleration in Conflict: Past and Present*. New York, NY: Cambridge University Press.

Gigerenzer, G., Krauss, S., and Vitouch, O. (2004). "The null ritual: what you always wanted to know about significance testing but were afraid to ask," in *The Sage Handbook of Quantitative Methodology for the Social Sciences*, ed D. Kaplan (Thousand Oaks, CA: Sage), 391–408. doi: 10.4135/9781412986311.n21

Honneth, A. (1995). *The Struggle for Recognition: The Moral Grammar of Social Conflicts*. Cambridge: Polity Press.

Lüdtke, O., Robitzsch, A., and Wagner, J. (2018). More stable estimation of the STARTS model: a Bayesian approach using Markov chain Monte Carlo techniques. *Psychol. Methods* 23, 570–593. doi: 10.1037/met0000155

Mayrhofer, R., and Hutmacher, F. (2020). The principle of inversion: why the quantitative-empirical paradigm cannot serve as a unifying basis for psychology as an academic discipline. *Front. Psychol.* 11:596425. doi: 10.3389/fpsyg.2020.596425

Muthén, B. O., and Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol. Methods* 17, 313–335. doi: 10.1037/a0026802

Nalborczyk, L., Bürkner, P. C., and Williams, D. R. (2019). Pragmatism should not be a substitute for statistical literacy, a commentary on Albers, Kiers, and van Ravenzwaaij (2018). *Collabra Psychol.* 5, 1–5. doi: 10.1525/collabra.197

Popper, K. (2013). *The Open Society and Its Enemies*. Princeton, NJ: Princeton University Press.

Rigdon, E. E., Sarstedt, M., and Ringle, C. M. (2017). On comparing results from CB-SEM and PLS-SEM. Five perspectives and five recommendations. *Market. Zfp* 39, 4–16. doi: 10.15358/0344-1369-2017-3-4

Scanlon, T. M. (2003). *The Difficulty of Tolerance: Essays in Political Philosophy*. New York, NY: Cambridge University Press.

Simon, B. (2020). A new perspective on intergroup conflict: the social psychology of politicized struggles for recognition. *Theory Psychol.* 30, 147–163. doi: 10.1177/0959354319887227

Simon, B., Eschert, S., Schaefer, C. D., Reininger, K. M., Zitzmann, S., and Smith, H. J. (2019). Disapproved, but tolerated: the role of respect in outgroup tolerance. *Pers. Soc. Psychol. Bull.* 45, 406–415. doi: 10.1177/0146167218787810

Simon, B., and Schaefer, C. D. (2016). Tolerance as a function of disapproval and respect: the case of Muslims. *Br. J. Soc. Psychol.* 55, 375–383. doi: 10.1111/bjso.12137

Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. M., and Lauro, C. (2005). PLS path modeling. *Comput. Stat. Data Anal.* 48, 159–205. doi: 10.1016/j.csda.2004.03.005

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., and Wetherell, M. (1987). *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Blackwell.

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., and Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: the last 25 years. *Psychol. Methods* 22, 217–239. doi: 10.1037/met0000100

Verkuyten, M., Adelman, L., and Yogeeswaran, K. (2020). The psychology of intolerance: unpacking diverse understandings of intolerance. *Curr. Direct. Psychol. Sci.* 29, 467–472. doi: 10.1177/0963721420924763

Wold, H. O. A. (1982). "Soft modeling: the basic design and some extensions," in *Systems Under Indirect Observations. Causality – Structure – Prediction: Part II*, eds K. G. Jöreskog and H. O. A. Wold (Amsterdam: North-Holland), 1–54.

Zitzmann, S., and Helm, C. (2021). Multilevel analysis of mediation, moderation, and nonlinear effects in small samples, using expected a posteriori estimates of factor scores. *Struct. Equat. Model.* doi: 10.1080/10705511.2020.1855076. [Epub ahead of print].

Zitzmann, S., Helm, C., and Hecht, M. (2021). Prior specification for more stable Bayesian estimation of multilevel latent variable models in small samples: a comparative investigation of two different approaches. *Front. Psychol.* 11:611267. doi: 10.3389/fpsyg.2020.611267

Zitzmann, S., Lüdtke, O., Robitzsch, A., and Hecht, M. (2020). On the performance of Bayesian approaches in small samples: a comment on Smid, McNeish, Miočević, and van de Schoot (2020). *Struct. Equat. Model.* doi: 10.1080/10705511.2020.1752216. [Epub ahead of print].

Conflict of Interest: The authors declare tha the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

REVIEW
published: 13 June 2022
doi: 10.3389/fpsyg.2022.710815

Dialectical Thinking: A Proposed Foundation for a Post-modern Psychology

Nikolay Veraksa[1], Michael Basseches[2] and Angela Brandão[3]*

[1]Faculty of Psychology, Lomonosov Moscow State University, Moscow, Russia, [2]Department of Psychology, Suffolk University, Boston, MA, United States, [3]Center for Philosophy of Sciences of the University of Lisbon, Lisbon, Portugal

For the authors, the way from a modern to a post-modern psychology requires dialectical thinking. Dialectical thinking recognizes the importance of contradiction, change, and synthesis; it also includes recognition of the value as well as limits of modern epistemological approaches. The article describes foundations for both ongoing efforts to understand and research the ontogeny of dialectical thinking and for appreciating the scope of dialectical thinking and its relevance for establishing a bridge from modern to post-modern psychology.

Keywords: dialectical thinking, post-modern psychology, limits of universalistic-formal analyses, limits of relativistic analyses, psychotherapy integration, ontogeny

OPEN ACCESS

Edited by:
Jana Uher,
University of Greenwich,
United Kingdom

Reviewed by:
Aaro Toomela,
Tallinn University, Estonia
Irina Mironenko,
Saint Petersburg State University, Russia

****Correspondence:***
Nikolay Veraksa
neveraksa@gmail.com

Specialty section:
This article was submitted to Theoretical and Philosophical Psychology, a section of the journal Frontiers in Psychology

Received: *17 May 2021*
Accepted: *14 March 2022*
Published: *13 June 2022*

Citation:
Veraksa N, Basseches M and Brandão A (2022) Dialectical Thinking: A Proposed Foundation for a Post-modern Psychology. Front. Psychol. 13:710815. doi: 10.3389/fpsyg.2022.710815

INTRODUCTION

The meanings of "modern" vs. "post-modern" is itself a topic for which several generations have contributed to an extensive literature. While space does not allow review of this literature, we could take as a jumping off point what could be considered a "popular" understanding as reflected on the internet "question-answering" website, Quora:

> The essential **difference between modern and postmodern** [in any discipline or field] is that modernism [generally the period 1890s to 1945] reflects rational thought and logic whereas **postmodernism** [postwar, nebulous start date] rejects logic. In short, modernism is theoretical and objective; postmodernism is subjective (Quora, 2014, first answer in Google.com search for "modern vs. postmodern").

From the authors' perspective, this quote reflects a popular misunderstanding that what postmodernism rejects is rational thought and logic in general, whereas, we understand it to reject a particular and very limited conception of rational thought and logic that is indeed associated with the concept of objectivity.

We understand the *modern perspective* on inquiry as assuming its purpose is to make sense of the world, we experience by discovering (or, in a more sophisticated version, creating, models of fundamental fixed realities)—basic elements and immutable laws. The rational thought and logic that *postmodernism* rejects is the specific logic of inquiry that assumes that such basic elements and immutable laws objectively exist, prior to and independent of the inquiry that describes them.

A post-modern psychology may begin by acknowledging the impact of the subjectivity of inquirers. It also asserts that the assumption that basic elements and immutable laws exist prior to and independent of the inquiry is a premise of modernist inquiry, rather than a justifiable ontological or epistemological principle. But in our view, a post-modern psychology must proceed from there to an articulation of a view of rational thought and logic that:

(a) makes epistemological progress possible while taking into account the role of *subjectivity in processes of inquiry*, and,
(b) includes the possibility of studying:

    (1) elements—as *changeable* and given meaning by the wholes/contexts of which they are parts, rather than as being fixed in nature and,
    (2) laws and regularities—as forms of organization that organize relationships among elements and that *change and develop over time*.

Without such an expanded view of the rationality that offers a model for the construction of more intersubjective, epistemically adequate understandings over time, post-modern inquiry would be limited to the simple accumulations of descriptions of various subjective interpretations of phenomena.

Post-modern critique in the absence of articulation of a more complete and powerful view of rationality and logic leads to what we would call "relativism." We view relativism as equally limited and epistemologically inadequate to the modernist approach of taking discovery of basic elements and immutable laws as the central goals of inquiry. Basseches' book on dialectical thinking (1984), while documenting the development of the capacity for dialectical thinking in adulthood, also made the case that a dialectical understanding of the nature of inquiry, rationality, and logic may be understood as providing a more epistemologically adequate organization of inquiry than either modernism or relativism taken by itself, while drawing heavily on the contributions of both modernists and post-modernists in various fields, including psychology.

In the psychological study of dialectical thinking three traditions can be differentiated. For convenience, we refer to them in a somewhat oversimplified way as "Russian," "Neo-Piagetian," (e.g., United States, Canada, and Western Europe), and "Asian" (e.g., China, India, and Japan). The Russian and Neo-Piagetian approaches focus more on the ontogeny of dialectical thinking, while the Asian approach focuses more on tendencies associated with culture. In this article's, overview of the development of dialectical thinking and its implications for psychology research and practice, we cite work in the Neo-Piagetian and Russian traditions. Here, we very briefly discuss the Asian approach, and why the article focuses on the other two.

The Asian approach is reflected in works by Nisbett, Peng (Peng and Nisbett, 1999) and Hamamura (Hamamura et al., 2008), among others. Peng et al. (2006) have studied a form of Eastern thought that they call naïve dialecticism and trace to East Asian Confucianism. It is characterized by the acceptance of contradiction and the idea of a constant flow or change in every aspect of the world, considering reality as a process (Spencer-Rodgers et al., 2010). This naïve dialecticism is characterized by a way of thinking that, contrary to Aristotelian logic, accepts oppositions and searches for a "middle way" or a compromise approach (Peng and Nisbett, 1999).

Account of naive dialecticism of Peng and Nisbett (1999) has been criticized by other researchers of Asian thought including Chan (2000) and Ho (2000). Ho criticized Peng and Nisbett for misunderstanding the relationships of contradiction with formal logic and dialectical thinking. Ho argues that formal logic applies to concrete problem analysis in a given moment, while dialectical thinking analyzes processes. Ho further states that Peng and Nisbett assume dialectical thinking is the cognitive ability to accept contradiction along with the "Confucian" non-dialectical tendency of Chinese to look for a compromise as a problem solution. Ho, following perspective of Basseches (1984), considers dialectical thinking as including metacognition that leads to revealing hidden implicit contradictions, striving to resolve them, and bringing the thinker to a more complex level of understanding of the problem.

We agree that Ho makes an important distinction and we propose the following points: First, naive dialecticism mainly represents a cultural style of thinking with its advantages and disadvantages, characterized by tolerance of contradictory beliefs (Hamamura et al., 2008; Ng and Hynie, 2016; Zheng, et al., 2021). Second, what has been called naive dialecticism can be described as a *dialogical relativism*. Naive dialecticism considers conflicting perspectives as belonging to the same whole within a process of constant change (i.e., it is dialogical). It thereby goes beyond the simpler perspective of relativism as description of various subjective interpretations of phenomena. It does not, however, specify any metacognitive capacities for discovering how to overcome contradictions thereby integrating the oppositions and creating new knowledge. Third, how to understand naive dialecticism in relation to the roots and forms of dialectical thinking found even in preschoolers by Russian psychologists (Shiyan, 2008; Bayanova, 2013; Veraksa et al., 2013), as well the "post-formal" organization of dialectical thinking studied by Basseches (1984) and his neo-Piagetian successors, are important questions for future study.

The work of each of the first two authors constitute contributions to understanding dialectical thinking as both a psychological phenomenon with a history in life-span human development, and as a systematic approach to inquiry that is applicable to a post-modern psychology. These two contributions can also be understood as extensions of two projects of the modern era which took shape in the early 20th century: (1) Jean Piaget's effort to create a genetic epistemology by modeling and illustrating the basic developmental processes that underlie the creation of organizations of understanding and (2) Lev Vygotsky's development of the dialectical point of view on psychological processes in the context of the historical tradition of dialectical analysis that took shape in the early 20th century in Russia. Both of these projects led to studies of dialectical thinking as a psychological phenomenon beginning in the

1970's and 1980's. We will summarize this work, and its relevance to establishment of a foundation for the development of a post-modern psychology.

DIALECTICAL THINKING IN PIAGET'S WORK AND NEO-PIAGETIAN RESEARCH ON DIALECTICAL THINKING AS A PSYCHOLOGICAL PHENOMENON

The genetic psychology of Piaget (1952, 1954, 1970, 1967) used psychological research to address epistemological questions. It studies and describes both structure and process in cognitive ontogenesis. The structural aspects include descriptions of forms of cognitive organization corresponding to the well-known stages of sensorimotor structures, representational structures, concrete operational structures, and formal operational structures. In each case, the organizational basis of the structure's stability is described, often in largely mathematical terms, and the features specified that make the structure more epistemically adequate than the ontogenetically prior organization—i.e., that enable the structure to deal with more complexity in more stable ways. But a clear tension can be identified between these *structural* aspects of Piaget's theory and his *framework for conceptualizing the process of development* of cognition and knowledge. In the *process* theory, consecutive moments of assimilation of novel experience to existing structures, and accommodation of those structures to novel experience lead to "conflict" or "disequilibrium, which requires "equilibration"—the construction of new more integrative forms of organization, to resolve. Piaget's *process* theory would surely be considered to be based on dialectical thinking using Basseches' criteria (1978, 1984). "The cognitive structures *described by* Piaget's theory could never account for the intellectual tools Piaget relied on *to create his theory*" is one way of expressing this tension. In this section, we will begin by discussing how Basseches (1984) addressed this tension.

Basseches (1978, 1984) began with an effort to describe dialectical thinking as an entirely different *form of organization of thought* that contrasted with the closed system form of organization of thought represented by Inhelder and Piaget's description of formal operational thought (1958). At the same time, Basseches advanced the argument that dialectical thinking, employing a model of open systems, interacting with, changing, and potentially becoming integrated with each other over time, provided a greater degree of equilibrium (a fundamental criterion of epistemological adequacy for Piaget) than closed-system modeling taken by itself. This greater equilibrium could explain movements from formal to dialectical thinking in adults as they encountered limitations of or contradictions among closed-system models, as well as offering a more adequate foundation for systematic inquiry. Basseches had the intention of describing a dialectical form of organization of thought that in neo-Piagetian terms could be described as "post-formal," implying hierarchical integration—i.e., that it could both make use of the power and value of formal analyses based on closed-system models, while at the same time being able to articulate the limits of formal analyses and transcend them. It could understand particular closed-system models as moments, within larger processes of contradiction and transformation—differentiation and integration. In address to the goals of this special issue, we are proposing that a post-modern psychology could similarly make use of analyses based on closed-system models, while at the same time using dialectical analyses to understand and transcend the limits of those closed-system analyses.

Thus, for understanding how dialectical analysis transcends the limits of closed-system analysis it is important to keep in mind the related Piagetian core concepts of *equilibrium* and *equilibration*. The cognitive concept of equilibrium, as the analogous biological concept of homeostasis, represents the capacity of the organism or system to maintain stability by adjusting to internal or external conditions that change the prior state. From the Piagetian genetic epistemological viewpoint, change represents development if, and only if, a reorganization of activity happened in such a way that, a new organization becomes capable of assimilating a greater variety of experiences, while maintaining stability in its core organizational features. The result of change in this case will be a cognitive organization with a *higher level of equilibrium*. Basseches (1984) argued that specific analyses as well as general approaches to inquiry based on the organizing principle of dialectic (which integrates dimensions of contradiction, change and system-transformation over time), will be more epistemically adequate than analyses and approaches to inquiry based on the formal operational principle of the closed system of lawful relationships of elements (Inhelder and Piaget, 1958).

Basseches (1984) addressed the following four questions, (1) "What is a dialectic?" (2) "How does a dialectical analysis, based on a model of dialectic, differ from a formal analysis, based on a model of a closed-system?" (3) "What gives dialectical analyses greater equilibrational/adaptive power" than formal analyses?" and (4) How can one identify, in examples of adult thought, the use of a dialectical model as well as formal models?"

Basseches' research used a sort of bootstrapping method. He started with what he recognized as dialectical analyses from intellectual history. He included a wide range of dialectical analyses of very different content that came from various intellectual disciplines with the intent of understanding *the underlying model of dialectic that they shared in common*.

Common features of these analyses were used both to derive a definition of dialectic and to recognize common patterns of underlying models as well as schemata or "moves-in-thought" that constituted "family resemblances" across this wide range of dialectical analyses that could also be recognized within the intellectual development of adults (Basseches, 1980, 1984). In his review of analyses from intellectual history, Basseches also recognized "dialectical" approaches, as well as "universalistic formal" and "relativistic" ones, as representing three alternative sets of "styles of inquiry," "intellectual sensibilities," and "world outlooks." The latter two sensibilities seemed to understand the goals of inquiry in radically different

ways, while the dialectical sensibility understood the goals of inquiry in a third way, which both incorporated aspects of the universalistic and relativistic approaches and yet transcended both.[1]

Three Styles of Inquiry: Universalistic Formal, Relativistic, and Dialectical

Basseches (1984) characterized the *universalist formal approach* to inquiry by the assumptions that there is a "universal order to things" which is the foundation for establishing "fixed universal truths." This order lends itself to description in "an abstract and formal way" and "all manner of phenomena in the universe may be found to fit in their places within this order." Systematic inquiry, whether scientific or philosophical, is aimed at describing this order. The sentiments associated with this approach are positive ones toward powerful abstract systems of ordering which capture the commonality or relationships among apparently different things." A good example of a system which was greeted with great enthusiasm is Chomsky's (1957) work in linguistics. Chomsky described linguistic structures which he claimed to be at the core of all the languages of human speakers all over the world, regardless of the many phenotypical differences among human languages. Universalistic formalists tend to have negative sentiments toward relativistic reasoning, which they view as "accepting too much sloppiness or disorder in the workings of the universe." They sometimes even view relativistic reasoning "as sloppy thinking which has retreated from the task of imposing strict order on everything" with an implication that this retreat might be due to "laziness" or "lack of intellectual power" (Basseches, 1984, p. 10).

In contrast, Basseches wrote that the *relativistic approach* to inquiry assumes (1) "that there is not one universal order to things but … many orders…" and (2) "that different individuals, groups, or cultures order reality in different and incompatible ways. Thus, order in the universe is entirely relative to the people doing the ordering." Systematic inquiry, whether scientific or philosophical, is aimed at description, appreciation, and even creation of "as wide a range of different orderings as may exist and be interesting and useful." Appreciation of diversity is a common positive sentiment among relativists. They greet with enthusiasm work that shows how things can be looked at in a variety of ways. Two such examples are "anthropologists' ethnographies of distant cultures (e.g., Mead, 1928)" and "idiographic approaches in personality psychology" (e.g., Allport, 1937). Relativists also value "tolerance or mutual appreciation among people who order the universe in different ways." Relativists tend to have negative sentiments toward "what they perceive as imperialism, including intellectual imperialism." It is often seen as imperialistic (1) "when universalistic formalists claim that one way of ordering things is *the right way*, equally applicable to phenomena experienced by all persons, groups, and cultures," or (2) "when universalists create schemes which acknowledge diversity of orderings but then order these diversities themselves within some over-arching framework that imputes greater value to some orderings than others." Anthropological theories which treat some cultures' modes of ordering as "primitive," or personality psychologists who treat some individuals as "pathological," using "standards taken from the anthropologists' own 'civilized' culture" or the mental health community's culturally shared consensus of what is "healthy" respectively, are frequently seen as equally imperialistic to viewing one framework as universally valid. In sum, any view which claims that one person's way of viewing things is truer or better than another's is regarded with distrust if not hostility by relativists except perhaps relativists' own view that their way of "conducting inquiry is better than that of the universalists" (Basseches, 1984, p. 10–11).

Basseches characterized the *dialectical approach* as charting a third alternative course. This approach assumes that "the evolution of order in the universe is an ongoing process" and that the *process of finding and creating order* in the universe is fundamental to human life and inquiry. "Thus, systematic inquiry, whether scientific or philosophical, is aimed at contributing to this process, and it is the process itself for which dialectical thinkers' most positive sentiments are reserved." Therefore, dialectical thinkers tend to "regard positively that which contributes to these processes and negatively that which obstructs them." The process of creating order is understood "as occurring through efforts to discover what is left out of existing ways of ordering the universe, and then to create new orderings which embrace and include what was previously excluded." Basseches continued:

> Dialectical thinkers can therefore be expected to share with universalistic formalists the negative reaction to relativistic reasoning, when the latter seems simply to acknowledge difference and disorder, and to retreat from efforts to find and create more powerful orderings. At the same time dialectical thinkers would share with relativists the reaction that it is dangerous to believe that an all-inclusive ordering is possible. For it is precisely when one thinks one has a achieved such an ordering that one stops actively looking for what is left out and what is different and in fact, (universalistic formalists) start to systematically defend (themselves) against perceiving such phenomena, (at which point inquiry may become limited to the extension of previous structures of ordering to newly investigated phenomena). [Basseches, 1984, p. 11].

Basseches goes on to comment on the issue of imperialism from the perspective of dialectical inquiry:

> Imperialism forces a way of life on others making it less likely that their own preferred way of life will be expressed. Intellectual imperialism imposes an order on the lives and meanings of others, making it less likely that the orderings created by others will be perceived. The easing up on the quest to find difference and disorder disrupts the fundamental process of inquiry as much as does the easing up on the effort to try to create

[1]See (Basseches, 1984, p. 10–12).

> order and unity when disorder and differences are discovered (Basseches, 1984, p. 11).

To end his characterization of a dialectical orientation toward inquiry, Basseches added that both types of work most appreciated by universalists and relativists are also valued by dialectical thinkers. However, the core of the dialectical approach is the locating of both types of work within an ongoing process of differentiation and integration of knowledge which transcends the limits of each approach to inquiry, taken by itself. We note that this very way of characterizing the three approaches represents a differentiation of universalistic formal from relativistic approaches to inquiry, and that the equilibrative power of dialectical thinking and dialectical analyses is demonstrated in the integration of the two previously differentiated approaches. Our argument in this paper is one for similarly differentiating modern approaches to psychology from approaches of post-modern psychology that appear to trade the acceptance of the role of subjectivity for a rejection of rationality, and then basing the further development of post-modern psychology on an integration of these two differentiated components. Later in this article, we will consider the nature of dialectical analyses, and the way they represent products of a dialectical approach to inquiry.

Returning to Basseches' research, the next step after a description of the implicit model that organizes dialectical thinking and the component schemata by which dialectical thinking could be recognized was as follows. Pilot interviews were conducted with individuals about matters important to them to see if the previously identified schemata and underlying models derived from intellectual history could be found in the spontaneous thinking of individual adults, thereby identifying dialectical thinking as a psychological phenomenon. In these interviews, an adaptation of Piagetian "clinical method" (Flavell, 1967, p. 47) was used to probe the limits of participants' capacities for cognitive organization. Based on these pilot interviews, the idea of dialectic was further clarified and the list of dialectical schemata recognizable in interviews was expanded. In the final phase, open-ended interviews related to the topic of the nature of education with the same structure of questions and the same "clinical method" interview techniques of probing were conducted with nine first-year students, nine fourth year students, and nine faculty members of the same highly selective liberal arts college. On the basis of these interviews, a measure of the development of the organizational form of dialectical thinking in an individual was developed and used to compare the cognitive approaches of members of all three subgroups.

Basseches (1984, p. 22) offered the following definition in answer to his first question, "What is (a) dialectic?": "*Dialectic is developmental transformation* (i.e., *developmental movement through forms*) *which occurs via constitutive and interactive relationships*."

To illustrate the components of this definition, we consider the forms of organization of theory and research on reproductive processes of pupfish (*Cyprinodon pecosensis*; Kodric-Brown, 1986). The phrase in the definition, "movement *through* forms" is meant to distinguish such movement from movement *within* forms. According to this research, when there is ample territory for nesting space habitat and substrate conditions are good, the reproductive activity is organized by dominance hierarchy. Thus, the activities of individual fish can be understood as movement *within* the form of the dominance hierarchy. As a reproductive season approaches, the mature males, who develop a bright blue color, engage in activities which establish their places in a dominance hierarchy. Then, in order of the hierarchy from top to bottom, each male chooses a "territory" which he tries to make as attractive as possible to welcome female pupfish to lay eggs. The females, who have a much more neutral color, in an order mainly organized by size (largest first), explore males' territories and choose males' territories in which to lay eggs. Males fertilize the eggs in their territories where they are protected until the baby fish are born.

In contrast, a movement *through* forms—a dialectic, occurs when due to some combination of the activity of the pupfish and other influences on their environment (seen dialectically as in *constitutive and interactive relationships*—to be elaborated below), a scarcity of territory suitable for nesting arises. This leads to the emergence of a more differentiated form of social organization in which males divide into three types, differentiated by reproductive strategy and appearance. The larger males become highly territorial, trying to create and defend their territories, while the smaller males become either "satellites" or "sneak-spawners." The satellites are smaller in size than males with territories but have similar phenotypical appearances. They function as parasites on the territories, and they reproduce by disrupting and stalling copulation by the territorial males and managing to fertilize a few eggs themselves. In contrast, the sneak-spawners become more like females both behaviorally and in their phenotypical appearance, remaining closer to females in size and not having a blue color. To other males, they appear to be females passing through territories, while deciding where to lay eggs; but at the same time they take advantage of the opportunity to fertilize the eggs of the "true females" who have laid them. Thus the movement from a form of organization based on (1) sexual dimorphism, (2) similar reproductive strategies among males, and (3) dominance hierarchies, to one based on (1) "multimorphism," (2) differences among males' reproductive strategies, and (3) territoriality among only larger males, is an example of a dialectical movement *through* forms. To refer to this transformation as "developmental" implies that there is a direction to it—a direction associated with greater complexity and increased species-capacity for adaptation.

The definition of *dialectic* states that the development transformation occurs *via constitutive* and *interactive* relationships. The adjective "constitutive" means that the relationships play a role in the making the parties what they are. The adjective "interactive" implies that a relationship is not static but is characterized by motion or action of the parties upon each other. We may use the same example to illustrate these concepts. *Constitutive and interactive relationships* can be identified among the individual pupfish and between the group of pupfish and their environment. An individual pupfish's activity is organized or *constituted* by its relationships with other pupfish as is clear when we say that a pupfish acts in accordance with its role in a dominance hierarchy. Role in a dominance hierarchy is not a characteristic of the pupfish

as a separate entity, but a characteristic of that pupfish's relationship to other pupfish. When we characterize an environment as "providing ample territory and an adequate substrate" for the reproduction of all the pupfish in it, we are not describing the environment in terms that frame it as prior to or separate from its relationship with pupfish, but rather in terms that depend on or are *constituted* by its relationship with pupfish.

The relationships among the pupfish and between the pupfish and their environment is interactive, as well as constitutive, in that the behavior of other pupfish affect the reproductive approach/activity and relative reproductive success for any given pupfish. Also, the activity of pupfish (e.g., creating overpopulation) can make a formerly adequate reproductive environment no longer adequate, while changes in the environment brought about by other factors than the pupfish' activity, can also render the environment no longer adequate, leading to transformation of the social organization of the pupfish.

To summarize how the whole process described above provides an illustration of *dialectic*, the *interaction* within the *constitutive relationships* of pupfish and their environments generates a limitation to the viability of the earlier *form* of reproductive activity, which in turn results in a *developmental transformation* (*movement through forms*) to a new, more complex *form* of reproductive activity. Thus the entire process described above can be seen as an example of *dialectic*.

The next steps for Basseches were (1) to illustrate the differences between "dialectical analyses"—based on using the model of dialectic described above to understand all matter of phenomena, and "formal analyses"—based on using the model of a closed system of lawful relationships described by Piaget to understand phenomena, as well as (2) to show the greater equilibrative power of dialectical analyses. While formal analyses can be understood as efforts to identify and describe fundamental unchanging laws, dialectical analyses seek to identify and describe fundamental processes of change and the dynamic relationships through which such change occurs. Reflecting Basseches' intent to derive from intellectual history an approach to understanding cognitive development in the lives of individual adults, he provided illustrations of the nature and power of dialectical analyses in both spheres.

While acknowledging the potential utility of the formal analyses, he tried to illustrate the power of dialectical analyses by showing the boundary conditions to which the scope of the value of formal analyses is limited. We recommend a similar approach to the creation of a post-modern psychology—demonstrating how the value of products of modern psychology is limited in each case to specific boundary conditions and proposing dialectical analyses and efforts at integration as the process of transcending those limitations.

Regarding intellectual history, Basseches (1984) wrote:

> Dialectical analyses can be found in the history of a wide range of intellectual disciplines, representing the natural sciences (Provine, 1971; Feyerabend, 1975; Horz et al., 1980), social sciences (Jay, 1973; Mandel, 1973; Kilminster, 1979), and humanities (Jameson, 1971; Adorno and Horkheimer, 1979).[2] They have been used to support political stances ranging from the very conservative (Hegel, 1821/1965) to the revolutionary (Marx and Engels, 1848/1955).

Two examples of dialectical analyses that are relatively well-known across academic fields, are:

(1) Marx's analysis (1844/1967) of the history of human productive and reproductive activity as a dialectical process in which (a) many aspects of economic, social, technical, and intellectual life are all interrelated within a form of organization of modes and social relations of production, and (b) tensions develop within these interrelationships and ultimately lead to a new form of organization of productive and reproductive activity (e.g., the replacement of feudal organization by capitalist organization); and
(2) Kuhn's dialectical analysis (1970) of the history of science in which the central ideas are (a) that research is shaped by dominant paradigms, (b) while paradigms make assumptions which serve as foundational premises for research, central to paradigms are pieces of insight-yielding research which can serve as a model for other researchers to follow, (c) in following paradigms subsequent research produces "anomalies" which are not easily reconciled with other extant knowledge, which in turn create discomfort among scientists, (d) while some scientists create and try to support *ad hoc* theories that preserve the dominant paradigms within increasingly unwieldy organizations of knowledge, other scientists start to create alternative paradigms which compete for followers with the dominant paradigm, (e) when a new paradigm (including assumptions, methodology, and ways of defining research problems and research solutions) attracts enough followers to represent a new dominant paradigm and redefine the nature of the field, a scientific revolution can be said to occur, and (f) While "normal science" guided by a paradigm can be characterized as puzzle-solving (corresponding to movement within forms in the definition of dialecticitc), the creation of new paradigms can be characterized as revolutionary (corresponding to transformational movement in the definition of dialectic).

Formal analyses in classical economic theory and philosophy of science respectively formed the backdrop against which dialectical analyses of Marx (1844/1967) and Kuhn (1970) were introduced as alternatives. Such formal analyses were characterized by assuming universally applicable laws—laws of economic behavior in one case and rules of evidence for hypothesis-testing in the other. The constitution of economic laws by existing relations of production and the possibility

[2]Kegan (1982) cites Wells (1972) as documenting a transformation toward more dialectical approaches in nearly every social and natural science during the last 150 years.

of transformation to new modes of production in which economic behavior would not follow the articulated laws were beyond the boundaries of scope of these formal analyses in the case of classical economic theory (Smith, 1776/1937). The constitution of rules of evidence by paradigms dominant in particular scientific communities and possibilities of scientific revolutions in which new paradigms would bring new rules of evidence were beyond the boundaries of the scope of formal analyses provided by confirmationist (e.g., Reichenbach, 1938) and falsificationist (e.g., Popper, 1959) philosophies of science.

Basseches' goal of describing a post-formal organization of dialectical thinking, which integrates the abilities to use and yet transcend the limitations of formal analyses can be seen in the examples of dialectical analyses from intellectual history and adult development. Marxist economic theory integrates using classical theory to understand laws of economic behavior under capitalism with analyzing limitations and actual and possible transformation of those laws. Kuhnian analysis can use philosophical analyses to clarify rules of evidence which currently organize a discipline while simultaneously analyzing historically how current paradigms achieved hegemony and where they may confront their limits. Basseches provided several examples of typical challenges of adult life for which he contrasted the use of dialectical analysis with formal analysis. In these examples, he also considers "relativism" as an approach to analysis, as the relationships between relativistic thinking and formalistic and dialectical thinking respectively is another matter systematically considered in the book that is quite relevant to the topic of creation of a post-modern psychology.

The neo-Piagetian dimension of Basseches approach is reaffirmed when he juxtaposes psychological and epistemological perspectives after acknowledging that in questioning limitations and boundary conditions of individuals' and communities' assumptions, dialectical thinking trades off a degree of intellectual security for freedom from imposing limitations on self, communities, or outsiders. He asserts that "from the point of view of humanity, as an epistemic subject involved in an ongoing pursuit of truth, the added power made possible by the capacity for dialectical analysis seems important to recognize." He ends with the claim that "dialectical thinking is an important psychological phenomenon, and that the capacity of dialectical thinking is an epistemologically important psychological attribute" (Basseches, 1984, p. 30).

Part II of *Dialectical thinking and adult development* (Basseches, 1984) presents his empirical work which addressed the fourth question, "How can one identify, in examples of adult thought, the use of a dialectical model as well as formal models." Although dialectical thinking is defined by an assumed underlying model of dialectic, dialectical thinking is identified by instances of schemata, or patterned movements-in-thought which dialectical thinkers tend to make. Over the course of pilot research, Basseches described 24 such patterned movements-in-thought that could be identified in interviews. When clear instances of a sufficient proportion of these schemata were observed, including the most complex schemata, the inference was made that the subjects' thinking was organized by a model of dialectic.[3]

In sum, Piaget's description of the equilibration process, in which assimilation and accommodation led to moments of disequilibrium, which were then resolved through the creation of more stable forms of cognitive organization, was at the core of his association between equilibrative power and epistemic adequacy. This concept could be applied both to the ontogeny and phylogeny of intelligence. In both cases, encounters with the limitations of prior forms of organization are what leads to more integrative forms of organization within which the capacities of earlier forms of organization were retained. But Piaget only began to treat dialectical thinking as an object of investigation in his later years (Piaget, 1974, 1977). Basseches, also in the 1970's, began work to extend Piaget's developmental and epistemological theory by describing dialectical thinking as a form of organization of thought that provided greater equilibrium than the most developed cognitive structure that Piaget had described, "formal operations." He claimed that while formal operations thinking relied on the idea of closed systems of lawful relationships in modeling various phenomena, dialectical thinking, with its application of the concept of developmental transformation over time, could articulate the boundary conditions for the utility of every closed-system model, and the processes by which those stability-focused models could become integrated into historical models of developmental transformation over time and the expansion of intersubjectivity that could occur with time. This idea is one pillar of our claim that dialectical thinking must be central to the most epistemically adequate approach to psychological inquiry.

In comparison with Piaget, the work of Lev Vygotsky was more explicit about the importance of dialectic in psychological inquiry from its outset. But it was left to those "neo-Vygotskians" who followed him, including Nikolay Veraksa, to begin an inquiry into dialectical thinking as a developmental psychological phenomenon. We discuss some of that work in the next section.

RUSSIAN RESEARCH ON DIALECTICAL THINKING AS A PSYCHOLOGICAL PHENOMENON

Research on dialectical cognition in Russian psychology is closely connected to the work of Lev Vygotsky, who asserted himself as a dialectician and stated that "all true scientific thinking moves along the path of dialectics" (Vygotsky, 1983,

[3]The presentation of the 24 schemata abstracted from various writings reflecting dialectical world-outlooks, then supplemented and modified based on pilot interviews, is beyond the scope of this article. The readers that want to go deeper on this subject should read the book "dialectical thinking and adult development" by Michael Basseches. Based on his data set, Chapter 4 of the book describes the overall organization of the 24 dialectical schemata and provides several examples of each dialectical schema drawn from his interviews with research subjects.

p. 37). From a post-modern perspective, we wince at such broad claims to truth, and we recognize a variety of forms of scientific thinking and inquiry. But seeing neither modernist nor relativistic forms of inquiry as adequate for a post-modern psychology, we affirm the importance of understanding the alternative "path of dialectics" which Vygotsky followed.

Vygotsky's emphasis on dialectical psychology was reinforced and further differentiated during Soviet times. Russian philosophy produced two concepts of dialectical cognition, which formed the context for further studies. One of them was based on the interpretation of *dialectical logic as a process* of understanding and dealing with evolving content, and the other stemmed from the analysis of *dialectical logic as a system of special operations and forms*.

The study of dialectical cognition, based on the latter understanding of dialectical logic, was started in the 1980s (Veraksa, 1981). The main distinguishing feature of this system that differentiates it from modern logic is that while modern logic perceives material or conceptual objects as basic fixed elements that are lawfully related to each other, dialectical logic treats such objects as *transformable structures*. Dialectical interpretation of these structures sees them as comprising relationships of opposites and allows the possibility that changes in such "internal" relationships of opposites can lead to changes in the structure of any such object. Studying the operation and development of the tools of this system of dialectical logic formed the basis of a research area that came to be known in Russia as "structural-dialectical developmental psychology" (Veraksa, 2006; Bayanova, 2013).

The *dialectical logic as a special system of logical tools* approach and the *dialectical logic as a process of understanding and dealing with evolving content* approach differed from each other not only in the foci and areas of their research. They also differed in the way their research was practically applied.

The analysis of Vygotsky's works reveals that to a high extent it was dedicated to the development of a dialectical psychology and the application of a dialectical approach to the psychological problems of dialectical method. Yet he did not see it as viable to use the dialectical method directly adopted from philosophy:

> (…) no psychological system can directly dominate psychology as a science without the help of methodology, that is, without creating a general science. The only legitimate adaptation of Marxism to psychology would be the creation of a general psychology with its concepts formulated in direct relation to general dialectics; that is, the dialectics of psychology. Any application of Marxism to psychology that follows other paths will inevitably lead to scholastic, verbal constructions, to dissolving dialectics in questionnaires and tests, to reasoning about things on the bases of external, casual, secondary features, to losing any objective criteria, to trying to negate any historical trend in the development of psychology, to a terminological revolution, in short, to a coarse deformation of Marxism and psychology. (…) in need of an as yet undeveloped but inevitable theory of biological materialism and psychological materialism as an intermediate science, which explains the concrete application of the abstract theses of dialectical materialism to the given field of phenomena.
>
> Dialectics covers nature, thinking, history—it is the most general, maximally universal science. The theory of the psychological materialism or dialectics of psychology is what I called general psychology (Vygotsky, 1982, p. 419–420).

Vygotsky's words cited above prove that he indeed set a goal to create such a scientific psychological theory that would mediate general dialectics and psychology as an independent science. He clarified:

> (…) dialectics of psychology—this is what we may now call the general psychology (…) is the science of the most general forms of movement (in the form of behavior and knowledge of this movement), i.e., the dialectics of psychology is at the same time the dialectics of a man as the object of psychology, just as the dialectics of the natural sciences is at the same time the dialectics of nature (Vygotsky, 1982, p. 322).

From our point of view, Vygotsky at the same time (a) engages with the extant tensions in "content" that are essential parts of the evolution of that content while he (b) starts to represent the more "abstract" regularities in the dialectical form of such evolution across many different processes of content evolution. This allowed describing relatively stable moments in the development of any area of inquiry or any individual's cognition in terms of the oppositions integrated in those moments. It also allowed the construction of the space of opportunities for a developing entity and a certain anticipation of possible conflicts and transformations at relatively stable and unstable moments in the course of its development.

We view Vygotsky as a dialectical constructivist who applied juxtaposition also in the framework of the analysis of the history of psychology:

> The development of scientific ideas and views is accomplished dialectically. Opposite points of view on the same subject replace one another in the process of developing scientific knowledge, and a new theory is often not a direct continuation of the previous one, but its dialectical negation (Vygotsky, 1982, p. 201).

Rubinstein (1957); Davydov (1972), and Leontiev (1983), among other psychologists and philosophers, emphasized the importance of using a dialectical materialistic analysis in psychological research. The question of this possible application of the dialectical method implied the development of a concept of dialectical cognition, as an antithesis, in tension with traditional formal thinking. In Russian (Soviet) philosophy, this question was given much prominence in works of Ilyenkov (1974, 1979). In particular, in his work "Dialectical Logic,"

he claimed that dialectical logic differed from the formal analysis of phenomena, the main difference being that dialectical logic dealt with contradiction.

> The dialectical logic is indeed the opposite for the formal logic, but not in the sense that dialectical logic rejects all the conclusions of the formal logic…. The main goal of formal logic is to discover general logical forms, laws, and rules independently from their particular content.
>
> Meanwhile, dialectical logic when solving the problem of truth at large, cannot be distracted from the concrete content of concepts, judgements, and inferences throughout the whole process of cognitive thinking (Andreev, 1985, p. 152–154).

Thus, the delineation is quite clear: the distinctive feature of formal logic became the operation of mental forms that were abstracted from their content, while dialectical logic began to be construed to be the logic of handling contradictions in the developing content.

It was argued that if dialectical logic existed, then it needed to possess the same formalism as formal logic did. In this case, it was to have different system operations (Maltsev, 1964). This opened up a second possibility of interpreting dialectics as the use of a "special logic" that differed from formal logic, and this circumstance could serve as a basis of dialectical cognition.

That created yet another perspective of dialectics as a special dialectical formal logic, both abstracted from its content and different from the traditional formal logic. This vision could lay the foundation for the construction of a formal theory of dialectical cognition, but only under a certain condition. That would be the description of abstract dialectical forms without material content. Therefore, such a description seemed basically unfeasible. This issue could be solved only if some unique dialectical operations were discovered, different from the traditional formal operations. Another obstacle was that almost all authors of that time acknowledged that dialectics comprised not only the analysis of development but also the analysis of the *emergence of new syntheses* within that process. Therefore, formalization of dialectics implied the formalization of the description of the process of development and the emergence of new syntheses. That, in its turn, looked totally impossible.

Many dialecticians agreed with that conclusion (Kopnin, 1973; Ilyenkov, 1979; Porus, 1979; Andreev, 1985, etc.).

Ilyenkov in particular encouraged:

> [The use of conceptual dialectical categories] (…) not as terms or catchphrases, but as forms of thinking, as active logical forms of study of objective reality. And first of all, the *category of contradiction* in its strictly objective definitions, which, being reflected in scientific consciousness of people and time-proved throughout centuries of their practical use actually are *logical* definitions of that category—unlike the ones given in mathematical logic where a contradiction is a synonym for "zero truth," where it is a synonym of "misperception" and "lie." In regards to the formal derivation of some combinations ("conjunctions") of signs from other combinations, these definitions are true, but they have nothing to do with thinking. Therefore, they cannot be called *logical* definitions of this concept (Ilyenkov, 1979, p. 143; words between brackets are those of the current authors).

Thus, Russian philosophy produced two different concepts of dialectical cognition—*dialectical logic as a process* of handling the evolving content and *dialectical logic as a system* of special operations and forms. It was in the context of these different concepts that further studies were conducted.

The former concept can be referred to as more "substantive," as dialectical analysis cannot be separated from the substantive content that is evolving. The understanding of dialectical cognition as handling evolving content was further developed by Davydov (1972), who relied heavily on Ilyenkov's work. Davydov considered dialectical cognition as a kind of thinking that analyzed the development of an entity based on its inner contradiction. In order to unfold the thought process in each specific analysis, it was necessary to find the initial key relation (contradiction) that gave rise to the whole variety of content as it formed. It is clear that the specific content presupposed the presence of a unique initial relationship, for which it was impossible to create a productive formal abstract analysis that disregarded the meaningful context in which the relationship formed and developed.

Thus, the approach represented by Davydov recognizes dialectical thinking in dialectical analyses that begin with the articulation of a specific core, which is generated by the initial contradiction and rooted in the conditions of a specific historical context and moment. It also traces out its sequelae over time from that key initial moment. As to the understanding of the mental or cognitive processes and capacities underlying an individual's creating of such an analysis, two specific features of dialectical cognition should be considered: (1) an all-round vision of the development of reality in its motion, and (2) avoiding abstracting from the content.

The study of dialectical cognition, based on the understanding of dialectical logic as a special kind of formal logic, was started in the 1980s (Veraksa, 1981). One of the problems to be solved in the course of these studies was to show the presence of formal dialectical operations that differed from those of formal logic. Since dialectical operations were to represent elements constituting and being constituted by a logical structure, the structure would need to be describable in mathematical terms in the same manner as it was done in the works of Jean Piaget.

The problem was solved in two stages. The first stage was associated with a search for some abstract units to be used for the analysis of dialectical cognition. Relations of opposition were chosen as such. They were construed as including any content or fragments of content that could be in opposition to one another and be in a relationship of mutual exclusion. Coming up with the idea of variability of the relationship where the opposites can find themselves in different situations

became the major breakthrough in the construction of dialectical logic and the understanding of dialectical cognition.

Veraksa and his colleagues found various examples in which opposites could be in different relationships: those of *transformation* (when one opposite is transformed into another); those of *transition* (when the transformation of one opposite into another does not occur immediately, but gradually, through an intermediate link); those of *conversion* (when the opposites pass into each other and then back); those of *mediation* (when for the two opposites are placed in such a situation where they act as components of the whole), etc (Veraksa, 2006).

The presence of such relationships between opposites apart from the mutual concordance and mutual exclusion allowed different consideration of the processes of object transformations. That required in the first place an identification of the characteristics that could be interpreted as opposites. Once they were ascertained, they could be matched with the schemes of previously discovered relationships which actually represented the options for possible transformations. In this case cognitive process unfolded on two levels: the formal abstract one (as handling the opposites) and on a concrete conceptual one (transition from the opposites to the content behind them).

Transformations described in such fashion could represent both the *real transformations of material objects* and *mental transformations of conceptual objects*. The former could be understood as processes occurring in inanimate and animate nature and the latter as dialectical mental acts. Obtained results in the form of descriptions of transformations of real and mental objects allowed linking them into a single logical structure.

The second stage was focused on the construction of a mathematical model of dialectical transformations (Veraksa and Zadadaev, 2012). Dialectical transformations were united into a dialectical structure D_n, a structure that was built using discrete mathematical tools. D_2 category became its elementary variant and was called a dialectical cycle. This cycle as a structure of dialectical logic has two extreme states that are in a relationship of opposition, and two opposite mediating states. If we denote the opposites through A and B, and their intermediate mediating states through AB and BA, then the simplest fragment of the dialectical structure can be illustrated with the help of **Figure 1**.

In this figure, the arrows denote the relationships of opposition and mediation. *The dialectical cycle differs from the traditionally understood dialectical triad of Thesis-Antithesis-Synthesis*. Of course, the dialectical cycle could be reduced by consecutive construction of triadic relations. However, there exist such objects and concepts that cannot be adequately described but with the help of a dialectical cyclic structure.

The structural-dialectical perspective sees the process of dialectical cognition as understanding of the transformations of the initial situation. First these transformations are represented in the conceptual plane, and then they appear as dialectical mental actions handling the relations of oppositions established by the subject for this particular situation (Veraksa, 2006). In connection with the above, *dialectical cognition is understood as a solution to a dialectical problem*. A dialectical task determines mental transformations performed by the subject.

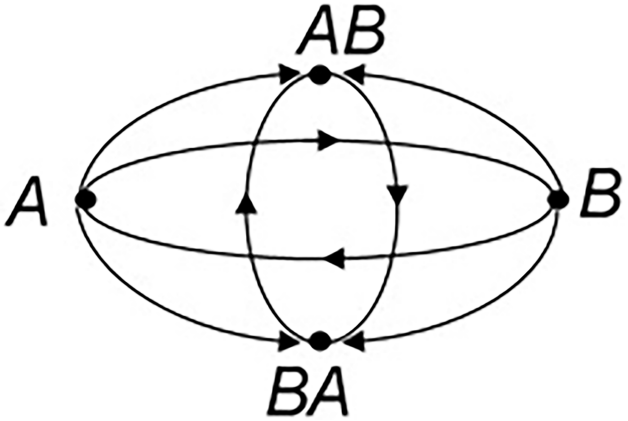


FIGURE 1 | The structure of dialectical cycle D_2.

First and foremost, we distinguish the action of a dialectical transformation among other dialectical mental actions. Its goal is to consider an object as its opposite. For example, Vygotsky (1983, p. 41), applied this dialectical action to the problem of defect: "A defect is not only weakness, but also strength. In this psychological truth is the alpha and omega of social education of children with disabilities."

Discovering the opposites and their relationship by means of dialectical cognitive actions became the foundation of structural-dialectical analytical method. Application of this method proves that an individual is indeed using the apparatus of dialectical thinking, being able to find the opposites and their relationship. In regard to the understanding of mental or cognitive processes behind this analysis, they are represented by mathematical models in the **Figure 1**. These models can be interpreted as a closed system described from the operational perspective. However, the dialectical structure depicted there contains development moving toward increasingly complicated levels. It is linked to the construction of new levels that are being created under a similar principle (Veraksa and Zadadaev, 2017).

One could assume that the structural-dialectical approach might not be efficient. Such a conclusion would be justified if application of the approach could be reduced to the statement of the fact of existence of the relations of opposites when characterizing various objects. Yet, the main feature of this method is actually the ability to describe the transformations of material and conceptual objects in relation to the opposites that the individual identified as the structural units. In other words, the structural-dialectical method allows comprehension of the logic of possible transformations or changes of an object and consideration of dialectical logic as the logic of opportunities. This is what constitutes the fundamental difference of structural interpretation of dialectical logic from traditional approaches. A new tool emerges, and it can serve not only for the registration of some occurring events but also for their anticipation. It formed the basis of a research area that came to be known in Russia as "structural-dialectical developmental psychology" (Bayanova, 2013; Veraksa et al., 2013).

With respect to the practical application of the results of each paradigm and line of research, the profound understanding of dialectical cognition as handling evolving content allowed Davydov to build a system of developmental education. It was based on the idea of identifying the contradictions that generate developmental content. If in a traditional school, teaching was based on the transition from the specific to the formation of abstract concepts through gradual generalization, developmental education proposed the opposite way—from a generalized relationship (which characterizes the main contradiction) to the construction of specific meaningful generalizations. This form of practice, however, presents a certain complexity, which is associated with higher requirements for a pedagogical team in charge of developmental training.

Within the framework of the structural-dialectical approach an adult education technology was developed around the "positional learning" model. It is focused on removing the estrangement between a discipline under study and a student's personality by engaging any content within contexts that are defined by different positions students take: e.g., the perspective of a critic, a poet, an apologist, etc (Veraksa, 1994).

A research area of no less importance is the development of preschool and school education. Since dialectical relations are part of any school subject and activity, they create a fundamental possibility of distinguishing in the knowledge system the meta-subject content accessible at different ages (Krasheninnikov, 2012). According to multiple researches, a distinctive feature of dialectical education is that the structure of preschoolers' dialectical cognition continues to be actively used at school age and subsequent ages (Shiyan, 2008), rather than being suppressed in the service of avoiding and eliminating contradiction as is advocated in traditional education.

PSYCHOTHERAPY RESEARCH AS AN EXAMPLE

To illustrate the differences among psychological inquiries founded on universalistic, relativistic, and dialectical thinking, we consider psychotherapy as an area of practice, research, and psychological curriculum. *Schools-cum-approaches* to psychotherapy initially developed more or less as separate, independent communities in the 20th century, each teaching approaches to theory and practice founded on its own assumptions, with little compatibility among them and little concern with their compatibility. Then, following what researchers following dialectical approaches expect, encounters among separate communities occurred, creating conflicts or issues to be resolved among them. Then, researchers taking a more macroscopic view of the subject matter differentiated major approaches which they referred to as psychodynamic, behavioral, existential-humanistic, cognitive, and systemic. Most of these researchers adopted a universalistic approach, within which they asked the question "which theory is most valid and which practice is most effective?," with the assumption that a straightforward methodology could be used to gather evidence that would satisfactorily provide the answers to these questions. Meanwhile, others instead contributed to a proliferation of different psychotherapies derived from clarifying different variations within these categories as well as from combining components across categories. But a general universalist zeitgeist led the vast majority of developers of the various "types" of therapy with different names to believe that they would need to gather evidence that demonstrated their theoretical and technical superiority to all of the other variants. There was little effort to articulate the historical conditions that brought these thriving communities of practitioners into conflict.

As predicted by hypothesis of Rosenzweig (1936) that claimed the existence of implicit common factors among all the different psychotherapies, only small and nonsignificant differences were found in effectiveness studies of various extant therapeutic approaches (Luborsky, et al., 1999, 2006). Some researchers and practitioners claimed these studies implied that most important for psychotherapy success would be to do "psychotherapy well" (i.e., following adequately one psychotherapy approach will ensure success, regardless of which one chooses to follow). We see this as adopting a largely relativistic approach to psychotherapy practice, research, and curriculum. Further in line with the ascendance of relativistic thought in this area, some therapeutic approaches to psychotherapy (e.g., narrative, constructivist) were founded on explicitly relativistic assumptions (e.g., every psychotherapy client is unique, and every psychotherapy process is unique, so searching for commonalities or general laws of psychotherapy is pointless).

Others facing the common factors hypothesis and psychotherapy challenges, in what could be seen as a dialectical approach to inquiry, espoused the hope for processes of psychotherapy integration to make therapy as effective and efficient as possible, considering both common and theory/technique-specific factors. These integration attempts followed different avenues (Kozarić-Kovacić, 2008; Castonguay et al., 2015): theoretical integration (i.e., creating one differentiated and integrated theoretical approach integrating the perspectives of communities in conflict), technical eclecticism (i.e., therapists being trained to be willing and able to implement effective techniques and elements from different communities), assimilative integration (i.e., using a single theoretical model but integrating ideas and techniques from other communities to overcome limits of effectiveness of one's model), and common factors approaches (i.e., focusing on common factors or processes within all psychotherapy communities). These attempts to integrate various communities' perspectives into a single theory/approach that would unite a wide range of theoretical and technical components of psychotherapy practice resulted, ironically, in a further proliferation of integrative theories of psychotherapy during 1980's and 1990s' decades. Garfield and Bergin (1994) already documented the existence of over 400 varieties of psychotherapy approaches varying according to their theoretical model, format (i.e., individual, family, and group), brief or long therapies, mental disorder type, and all sorts of combination of different elements. Consequently, what could be seen as an attempt to achieve a synthesis by adopting a dialectical perspective, led in many cases to a return to a universalistic perspective of trying to find which integrative approaches would most

increase effectiveness and efficiency for all cases! In our view, while the advocacy of integration was admirable, it was limited by the failure of the approach to be fully dialectical. It treated psychotherapy debates as isolated rather than occurring in the context of socio-economic factors leading to, for example, financial and status competition among advocates of various approaches, the increased role of insurance companies as supporters and therefore gate-keepers of psychotherapy, the need of a more educated group of consumers for protection from exploitation, etc.

Nevertheless, the integrative movement has continued to expand, representing a dialectical attitude of differentiating approaches as useful within specific social and historical contexts, while also integrating approaches to transcend limits of context. Approaches within that movement ask therapists to be responsive and flexible to patients' needs, taking into account their differences in contexts, cultural lessons learned, and problems (Zarbo et al., 2016). Furthermore, therapists have to recognize that patients' needs evolve across time and situations; to understand the importance of maintaining and adapting therapeutic alliances across these evolutions over time; and to understand commonalities as well as differences in *developmental process* across psychotherapy cases, both with respect to not only clients' repertoires but also the alliances themselves among therapists and clients (Basseches and Mascolo, 2010). Viewing psychotherapy cases as developmental processes implies that clients' development should be fostered, tracked, and evaluated in a process that considers not only the clients' adaptive challenges but also the adaptive challenges of all the others whom the clients' actions affect. A dialectical approach to practice and theory clearly demands a psychotherapy curriculum that prepares students to think dialectically about the complex set of interrelated phenomena that psychotherapy comprises, including to recognize the conflicts and potential synthetic resolutions that exist among those phenomena.

But why is dialectical thinking really needed for a post-modern inquiry into psychotherapy? The research briefly summarized before can be puzzling. It seems this field is going in circles between searching for a single common approach that fits all (universalistic formal analysis), and failing on that attempt, to a relativistic perspective that affirms uniqueness of cases and diversity of approaches, while undermining any attempt to provide therapists with avenues to guidance in making the moment-by-moment choices that are needed and to provide the field with any pathways for improving overall interventions and outcomes. Several factors contribute to this "going in circles effect." One aspect is the failure to emphasize that while there are operative "psychotherapy delivery systems" in most cultures, there are also environmental systems outside those systems which interact with the psychotherapy delivery systems. Such extra-therapeutic systems include client support systems (Drisko, 2004), which may be more or less adequate, as well as economical and sociological systems which influence the contexts and possibilities for psychotherapy. A dialectical approach allows going beyond the dichotomy of nomothetic vs. ideographic approaches to psychotherapy. Development in psychotherapy can be viewed as a movement *through* forms occurring due to some combination of the conflicts and conflict resolution activity within the constitutive and interactive relationships of clients and therapists, as well as within each of the parties' interactive and constitutive relationships with their environments. Thus, every therapy relationship is a dialectic, and the model of dialectic can be used to track the patterns and challenges within that relationship. The fact that in any psychotherapy process we are dealing with constant interaction within the therapy context and outside of it, requires a dialectic analysis.

In practical terms, we need to maintain awareness of the boundary conditions to which the scope of the value of formal analyses in psychotherapy is limited and include the case-specific context in a way that idiosyncrasies that violate those boundary conditions can be part of the dialectical analyses of psychotherapies. As we have seen in psychotherapy research, when this awareness is not taken into account we end as Vygotsky (1982) forewarned: following "other paths will inevitably lead to scholastic, verbal constructions, to dissolving dialectics in questionnaires and tests, to reasoning about things on the bases of external, casual, secondary features, to losing any objective criteria, to trying to negate any historical trend in the development of psychology."

Above, we cited the book "*psychotherapy as a developmental process*" of Basseches and Mascolo (2010) in which a dialectical method for assessing micro-developments in therapy across all psychotherapeutic approaches was proposed. The proposal offers a useful framework in differentiating three fundamentally different types of resources that therapists' actions within the therapy relationship can provide to the therapeutic process, regardless of what guiding model therapists are following and what therapeutic techniques they are employing. The proposal also claims, and has supported this claim with case studies, that the three different types of resource must be integrated for a case of psychotherapy to be successful in some way. Each resource, in its own way, fosters the emergence and exploration of conflict, and the authors' present a method for tracking utterance by utterance in verbatim dialog the steps that either lead to successful resolutions of conflicts or that leave them unresolved. Unfortunately, this research method is very labor-intensive which limits its practical utility. However, its implied guideline for clinical practice is valuable: To be aware of developmental processes on a moment-to-moment basis, regardless of whether one is following one or many therapeutic and/or technical approaches. What is still needed from a dialectical approach to psychotherapy are tools for clinicians to use to look at their practice at all levels of therapy (short, medium, and long term) including emerging conflicts and resolutions as they develop therapeutic relationships and work with various common factors, techniques, types of clinical problems, and contextual factors. Advances with this proposal or further different proposals will have to ensure ways of integrating nomothetical proposals with a clear map of how to navigate the ideographic aspects of each case, not forgetting that the central aspect of therapy is the transformational process! However, this requires clinicians learning to think dialectically, for which we hope a post-modern psychology will provide a supportive context. Neither a modernist/universalistic nor a relativistic psychology can offer therapists and students of therapy such important tools.

DISCUSSION

Returning to our introduction, we stated our view that without an expanded view of rational inquiry that offers a model for the construction of more intersubjective, epistemically adequate understandings over time, post-modern inquiry would be limited to the simple accumulations of descriptions of various subjective interpretations of phenomena.

Regarding the form of rational inquiry that may be at the core of modernist inquiry, all of the work on dialectical thinking that we have described recognized that this form of rationality was insufficient by itself. The Vygotskian tradition started with the content of ideas, studied historically. It described in such studies the appearance of contradictory ideas, followed by the appearance of synthetic ideas that resolved the contradictions, only to themselves later be contradicted as natural and social environments changed. It was presumed that through instruction using such historical analysis of content, students would learn to "think dialectically." Veraksa's "structural-dialectical developmental psychology" started with the observation that throughout ontogeny, encounters with relations of opposition and contradiction occurred, and that cognitive operations on relationships of opposition and contradiction (recognizing contradiction and transformation) were as essential for coping with both the material world and the conceptual world as operations (such as those described by Piaget) for creating, recognizing, and maintaining stability. It recognized that most educational and other socializing systems tended to privilege the creation and maintenance of fixed order over the appreciation of contradiction and the development of the ability to recognize it and to deal with it in creative and transformative ways. This led to research on questions of how early in child development could the use of dialectical operations be recognized, and how the tendency to suppress the development of this kind of thinking in favor of thinking which identifies, creates and maintains order, could be counteracted educationally.

The focus of Basseches (1984) was on the differences between the most powerful system of creating order described by Inhelder and Piaget (1958)—formal operational thought, and the most complex form of dialectical thinking (now sometimes referred to as "scientific dialectics") that he found in a wide range of contents of intellectual history, including Piaget's own account of the process of ontogenesis of the "organ" of intelligence. Basseches saw the capacities for such thinking as rooted in different models or forms of organization that were constructed over the course of cognitive ontogenesis. He proposed that the model underlying formal operational thought (and perhaps "modernist" inquiry) was a model of a "closed-system of lawful relationships." He contrasted this with a proposed model underlying dialectical thinking of open self-transforming systems in interaction with each other. He claimed that dialectical thinking represented a form of thinking more developed than formal operational thinking because the idea of dialectics required the understanding of a system and represented a differentiation of the concept of a closed system from that which was beyond the limits or boundary conditions of a closed system. Dialectical thinking thus included the capacities to use both closed system and open system models. Basseches' proposal paralleled Piaget's argument for why each of his stages represented development beyond the previous one and why each previous stage was a necessary but not sufficient condition for the subsequent stage. It could therefore be seen both as a critique of the limits of Piaget's project and an extension of that project beyond those limits.

In our view, the study of post-formal dialectical thinking is of great importance because it includes the power of formal analyses, while at the same time having the power to transcend the limitations of closed-system analyses.[4] Describing and identifying such thinking has been the focus of the work of Basseches and his colleagues. Putting neo-Piagetian and neo-Vygotskian streams of research together, we can ask the following question: how does identifying and promoting the development of dialectical thinking in various periods of childhood[5] affect the processes and the likelihood of individuals becoming capable in adulthood of creating dialectical analyses, organized by the idea of dialectic, as articulated by Basseches? With sufficient resources a long-term longitudinal study could be conducted that would begin to answer such questions.

Thus, we propose that the study of dialectical thinking be an effort to study the development of dialectical thinking capacities based on (1) acknowledgment of the value of efforts to organize human actions and observations, when contradictions are inevitably discovered and encountered; while at the same time and (2) treating specific organizations created as moments in dialectical processes, and not as fixed unchangeable laws of nature or of human activity.

We articulated our view in our first paragraph that the transition from a modern psychology to an adequate post-modern psychology depends on dialectical thinking. We can end by expanding that articulation in the light of the foregoing. We see modernist inquiry as in some way analogous, if not equivalent, to inquiry aimed at discovering and articulating fixed, lawful regularities undergirded by the structures of formal operational thought. Consistent with Riegel's (1973) argument, to cling tightly to modernist theories and interpretations of data would be to contribute to the phenomenon he referred to as "alienated thinking" (or "alienated knowing"). On the other hand, to simply reject the products of modern inquiry would be analogous to the devaluing of moments of creating organizing structures when faced with contradictory, opposite, or conflicting actions and observations. This would deny both the value of forms of stability for human life and the importance of bringing together different subjectivities to create intersubjectivities. Doing so would not create an adequate replacement for modern science and inquiry. As an alternative, we propose a post-modern psychology that understands inquiry as composed of temporary moments which represent steps in

[4]More generally, we would say that within the dialectics of cognitive ontogeny, those moments of creating order out of contradiction, as well as the moments of recognition of contradictions that challenge existing orderings, are equally important and mutually dependent and complementary. And this is equally true for the development of knowledge at the phylogenetic level as well.

[5]While counteracting the alienating effect of education that focuses on apprehending structures—those objectified (perhaps even fetishized—oh those correct answers on tests that translate into grades on report cards feel good!) products of organizing contradictory experience—at the expense of encouraging the process of dealing creatively and constructively with contradiction.

ongoing processes. Some steps entail finding contradictions between structures and that which is outside of, other than, or unstable within, such structures. Other steps entail transforming understandings to more complex, differentiated and integrated ones by conceptualizing the relationship between what is well-organized within a structure, and that which lies beyond or stands against its organizing power. We hope not only to have outlined a path for future study of dialectical thinking, but also to have implied a pathway for the development of post-modern psychology.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

The research was supported by the RSF grant 19-18-00521-П.

REFERENCES

Adorno, T. W., and Horkheimer, M. (1979). *Dialectic of Enlightenment*. London: NLB.

Allport, G. W. (1937). *Personality: A psychological interpretation*. New York: Holt, Rinehart, and Winston.

Andreev, I. D. (1985). *Dialectical Logic*. Moscow: Visshaya Shkola.

Basseches, M. (1978). Beyond closed-system problem-solving: A study of metasystematic aspects of mature thought. Doctoral Dissertation, Harvard University, University Microfilms, Ann Arbor, MI, USA.

Basseches, M. (1980). Dialectical schemata: a framework for the empirical study of the development of dialectical thinking. *Hum. Dev.* 23, 400–421. doi: 10.1159/000272600

Basseches, M. (1984). *Dialectical Thinking and Adult Development*. Norwood, NJ, USA: Ablex.

Basseches, M., and Mascolo, M. (2010). *Psychotherapy as a Developmental Process*. New York: Routledge.

Bayanova, L. F. (2013). Vygotsky's hamlet: the dialectic method and personality psychology. *Psychol. Rus. State Art* 5, 35–42. doi: 10.11621/pir.2013.0103

Castonguay, L. G., Eubanks, C. F., Goldfried, M. R., Muran, J. C., and Lutz, W. (2015). Research on psychotherapy integration: building on the past, looking to the future. *Psychother. Res.* 25, 365–382. doi: 10.1080/10503307.2015.1014010

Chan, S. F. (2000). Formal logic and dialectical thinking are not incongruent. *Am. Psychol.* 55, 1063–1064. doi: 10.1037/0003-066X.55.9.1063

Chomsky, M. (1957). *Syntactic Structures*. The Hague: Mounton.

Davydov, V. V. (1972). *Means of Generalization in Education*. Moscow: Pedagogika.

Drisko, J. W. (2004). Common factors in psychotherapy outcome: meta-analytic findings and their implications for practice and research. *Fam. Soc.* 85, 81–90. doi: 10.1606/1044-3894.239

Feyerabend, P. (1975). *Against Method: An Outline of an Anarchistic Theory of Knowledge*. London: NLB.

Flavell, J. (1967). *Genetic Psychology of Jean Piaget*. Moscow: Prosveshenie.

Garfield, S., and Bergin, A. (1994). "Introduction and historical overview," in *Handbook of Psychotherapy and Behaviour Change*. eds. A. Bergin and S. Garfield (Chichester: Wiley), 3–18.

Hamamura, T., Heine, S. J., and Paulhus, D. L. (2008). Cultural differences in response styles: the role of dialectical thinking. *Personal. Individ. Differ.* 44, 932–942. doi: 10.1016/j.paid.2007.10.034

Hegel, G. W. F. (1821/1965). *The Philosophy of Right*. Chicago: Encyclopedia Britannica.

Ho, D. Y. F. (2000). Dialectical thinking: neither eastern nor western. *Am. Psychol.* 55, 1064–1065. doi: 10.1037/0003-066X.55.9.1064

Horz, H., Poltz, H., Parthey, H., Rosenberg, U., and Wessel, K. (1980). *Philosophical Problems in Physical Science*. Minneapolis, MN: Marxist Educational Press.

Ilyenkov, E. V. (1974). *Dialectical Logic*. Moscow: Politizdat.

Ilyenkov, E. V. (1979). *Dialectical Contradiction*. Moscow: Politizdat.

Inhelder, B., and Piaget, J. (1958). *The Growth of Logical Thinking From Childhood to Adolescence*. New York: Basic Books.

Jameson, F. (1971). *Marxism and form: 20th century dialectical theories of literature*. Princeton, NJ: Princeton University Press.

Jay, M. (1973). *The Dialectical Imagination*. Boston: Little Brown and Company.

Kegan, R. (1982). *The Evolving Self: Problem and Process in Human Development*. Cambridge, MA, USA: Harvard University Press.

Kilminster, R. (1979). *Praxis and Method: A Sociological Dialogue With Lukacs, Gramsky, and the Early Frankfurt School*. London: Routledge & Kegan Paul.

Kodric-Brown, A. (1986). Satellites and sneakers: opportunistic male breeding tactics in pupfish (*Cyprinodon pecosensis*). *Behav. Ecol. Sociobiol.* 19, 425–432. doi: 10.1007/BF00300545

Kopnin, P. V. (1973). *Dialectics as Logic and Theory of Mind*. Moscow: Nauka.

Kozarić-Kovacić, D. (2008). Integrative psychotherapy. *Psychiatr. Danub.* 20, 352–363. PMID: 18827763

Krasheninnikov, E. E. (2012). *Development of Cognitive Abilities of Preschoolers. For Work With Children 4–7 Years Old*. Moscow: Mozaika-Sintez.

Kuhn, T. S. (1970). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Leontiev, A. N. (1983). *Selected Psychological Works: In 2 vols. Vol. II*. Moscow: Pedagogy.

Luborsky, L., Diguer, L., Luborsky, E., and Schmidt, K. A. (1999). "The efficacy of dynamic versus other psychotherapies: is it true that "everyone has won and all must have prizes?"—an update," in *Psychotherapy Indications and Outcomes*. ed. D. S. Janowsky (Washington, DC: American Psychiatric Press), 3–22.

Luborsky, L., Rosenthal, R., Diguer, L., Andrusyna, T. P., Berman, J. S., Levitt, J. T., et al. (2006). The dodo bird verdict is alive and well-mostly. *Clin. Psychol. Sci. Pract.* 9, 2–12. doi: 10.1093/clipsy.9.1.2

Maltsev, V. I. (1964). *Essay on dialectical logic*. Moscow: Moscow University Press.

Mandel, E. (1973). *An Introduction to Marxist Economic Theory*. New York: Pathfinder.

Marx, K. (1844/1967). in *Economic and Philosophical Manuscripts in Writings of the Young Marx on Philosophy and Society*. eds. L. D. Easton and K. H. Guddat (Garden City, New York: Doubleday & Company)

Marx, K., and Engels, F. (1848/1955). *The Communist Manifesto*. New York: Appleton-Century Crofts.

Mead, M. (1928). *Coming of age in Samoa*. New York: Morrow.

Ng, A. H., and Hynie, M. (2016). Naïve dialecticism and indecisiveness: mediating mechanism and downstream consequences. *J. Cross-Cult. Psychol.* 47, 263–276. doi: 10.1177/0022022115613861

Peng, K., and Nisbett, R. E. (1999). Culture, dialectics, and reasoning about contradiction. *Am. Psychol.* 54, 741–754. doi: 10.1037/0003-066X.54.9.741

Peng, K., Spencer-Rodgers, J., and Nian, Z. (2006). "Naïve dialecticism and the Tao of Chinese thought," in *Indigenous and Cultural Psychology: Understanding People in Context*. eds. U. Kim, K.-S. Yang and K.-K. Hwang (Springer Science + Business Media).

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: Norton.

Piaget, J. (1954). *The Construction of Reality in the Child*. New York: Basic Books.

Piaget, J. (1967). *Six psychological studies*. New York: Random House.

Piaget, J. (1970). *Structuralism*. New York: Basic Books.

Piaget, J. (1974). *Recherches sur la Contradiction. Etudes D'Epistemologie Genetique.* Paris:P.U.F. *31*, 32.

Piaget, J. (1977). *Recherches Sur l'Abstraction Reflechissante. Etudes d'Epistemologie Genetique*. Paris:P.U.F. *34*, 35.

Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.

Porus, V. N. (1979). *Dialectical contradiction*. Moscow: Politizdat.

Provine, W. (1971). *The Origins of Theoretical Population Genetics*. Chicago, IL: University of Chicago Press.

Quora (2014). First answer in Google.com search for "modern vs. postmodern" by Roberta Pearce in Quora. Available at: https://www.quora.com/What-is-the-difference-between-modern-and-postmodern-art?top_ans=5280438

Reichenbach, H. (1938). *Experience and Prediction*. Chicago, IL: University of Chicago Press.

Riegel, K. F. (1973). Dialectic operations: the final period of cognitive development. *Hum. Dev.* 16, 346–370. doi: 10.1159/000271287

Rosenzweig, S. (1936). Some implicit common factors in diverse methods of psychotherapy. *Am. J. Orthop.* 6, 412–415. doi: 10.1111/j.1939-0025.1936.tb05248.x

Rubinstein, S. L. (1957). *Being and Consciousness*. Moscow: Izdatelstvo AN USSR.

Shiyan, O. A. (2008). Developing education in university: dialectical structure of the educational course as a source of student development. *Psychol. Sci. Educ.* 2, 9–17. doi: 10.17759/pse

Smith, A. (1776/1937). *An Inquiry Into the Nature and Causes of the Wealth of Nations*. New York: Modern Library.

Spencer-Rodgers, J., Williams, M. J., and Peng, K. (2010). Cultural differences in expectations of change and tolerance for contradiction: a decade of empirical research. *Personal. Soc. Psychol. Rev.* 14, 296–312. doi: 10.1177/1088868310362982

Veraksa, N. (1981). Peculiarities of the transformation of conflicting problem situations by preschool children. *Voptosy Psikhol.* 3, 123–127.

Veraksa, N. (1994). The model of standpoint teaching of students. *Voptosy Psikhol.* 3, 122–129.

Veraksa, N. (2006). *Dialectical Thinking*. Ufa: Vagant.

Veraksa, N., Belolutskaya, A., Vorobyeva, I., Krasheninnikov, E., Rachkova, E., Shiyan, I., et al. (2013). Structural dialectical approach in psychology: problems and research results. *Psychol. Rus. State Art* 6, 65–77. doi: 10.11621/pir.2013.0206

Veraksa, N., and Zadadaev, S. (2012). Dialectical thinking and W-measure of the development of two-dimensional dialectical structure. *Psychol. Sci.* 15, 57–86.

Veraksa, N., and Zadadaev, S. (2017). Foundations of structural dialectics. Herald of the MCU 3, 8–25.

Vygotsky, L. S. (1982). *Collected Works in 6 1*. Moscow: Pedagogika.

Vygotsky, L. S. (1983). *Collected Works in 6 5*. Moscow: Pedagogika.

Wells, H. K. (1972). Alienation and dialectical logic. *Kansas J. Sociol.* 3, 1.

Zarbo, C., Tasca, G. A., Cattafi, F., and Compare, A. (2016). Integrative psychotherapy works. *Front. Psychol.* 6:2021. doi: 10.3389/fpsyg.2015.02021

Zheng, W., Yu, A., Li, D., Fang, P., and Peng, K. (2021). Cultural differences in mixed emotions: the role of dialectical thinking. *Front. Psychol.* 11:538793. doi: 10.3389/fpsyg.2020.538793

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

PERSPECTIVE
published: 17 June 2022
doi: 10.3389/fpsyg.2022.908311

Future Objectivity Requires Perspective and Forward Combinatorial Meta-Analyses

*Barbara Hanfstingl**

Institute for School and Instructional Development, University of Klagenfurt, Klagenfurt, Austria

This manuscript contributes to a future definition of objectivity by bringing together recent statements in epistemology and methodology. It outlines how improved objectivity can be achieved by systematically incorporating multiple perspectives, thereby improving the validity of science. The more result-biasing perspectives are known, the more a phenomenon of interest can be disentangled from these perspectives. Approaches that call for the integration of perspective into objectivity at the epistemological level or that systematically incorporate different perspectives at the statistical level already exist and are brought together in the manuscript. Recent developments in research methodology, such as transparency, reproducibility of research processes, pre-registration of studies, or free access to raw data, analysis strategies, and syntax, promote the explication of perspectives because they make the entire research process visible. How the explication of perspectives can be done practically is outlined in the manuscript. As a result, future research programs can be organized in such a way that meta-analyses and meta-meta-analyses can be conducted not only backward but forward and prospectively as a regular and thus well-prepared part of objectification and validation processes.

Keywords: objectivity, perspective, subjectivity, specification curve analysis, meta–meta-analyses, combinatorial meta-analysis

OPEN ACCESS

Edited by:
Alexander Robitzsch, IPN - Leibniz Institute for Science and Mathematics Education, Germany

Reviewed by:
Oliver Luedtke, University of Kiel, Germany
Matthias Borgstede, University of Bamberg, Germany

***Correspondence:**
Barbara Hanfstingl
barbara.hanfstingl@aau.at

Specialty section:
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

Received: *30 March 2022*
Accepted: *01 June 2022*
Published: *17 June 2022*

Citation:
Hanfstingl B (2022) Future Objectivity Requires Perspective and Forward Combinatorial Meta-Analyses. Front. Psychol. 13:908311. doi: 10.3389/fpsyg.2022.908311

INTRODUCTION

Objectivity is a core criterion for achieving sound scientific results. It reflects a central specificity of modern science. The concept bears different, although related, definitions. According to Gaukroger (2012), objectivity is the capacity to stand back from our perceptions, beliefs, and opinions, and to shift perspective. From an anthropological perspective, Tomasello (2020) sees the need for social inclusion already in children, which requires objectivity as a developmental prerequisite for adopting different perspectives and social inclusion. The American Psychological Association (APA)[1] defines objectivity as (1) "the tendency to base judgments and interpretations on external data rather than on subjective factors, such as personal feelings, beliefs, and experiences; (2) a quality of a research study such that its hypotheses, choices of variables studied, measurements, techniques of control, and observations are as free from bias as possible;" and as opposite of subjectivity. In test statistics, objectivity is one of the three main quality criteria for psychological

[1]https://dictionary.apa.org/

tests, along with reliability and validity, and refers to the test's procedure, result, analysis, and interpretation, being independent of the person conducting the test.

The modern idea of objectivity grew in times of Enlightenment when scientific thinking took over the dogmatic thinking of the Christian church in Europe. According to Daston and Galison (2010), who provide a comprehensive history of the concept, objectivity, as we use it in science today, emerged in the mid-nineteenth century and is associated with the ability to display "the world as it is," aided by the invention of photography. From a traditional scientific view, Popper (1972) saw objectivity as the correct application of scientific methods and procedures to make accurate predictions. This view on objectivity comes very close to what recent methodologies understand as objectivity: findings are scientific if they are reproducible and intersubjectively testable. Scientific thinking deals with objective facts; if knowledge is scientific, it is objective and objectifiable. Ideally, objectivity ensures the necessary distance to the subject of interest, it prevents the scientist from subjectivity and (emotional) involvement.

Scientific communities agree that objectivity is necessary to ensure that a scientific fact is indeed a scientific fact. Nevertheless, several authors from different fields question the current concept of scientific objectivity as a sufficient criterion for establishing a scientific fact. In the Stanford Encyclopedia of Philosophy, Reiss and Sprenger (2020) question the reachability of objectivity and see the final understanding of it as an ongoing project. As early as 1933, Rosenthal and Rosnow (2009; reprint) point out in their research the influences of human bias – subjectivity, to put it succinctly – on research findings. Recent state-of-the-art publications show that human bias significantly affects scientific results, even when we strive for objectivity in testing the same hypotheses and use accepted scientific methods to test them (e.g., Silberzahn et al., 2018; Bastiaansen et al., 2020; Schweinsberg et al., 2021). These studies show that the current state of the methodology does not meet objectivity, as researchers must make individual decisions and specifications on how to conduct a study. Therefore, one reason we struggle with objectivity lies in the historical and social developments of the so-called postmodern era. Both the world and science are realized as increasingly complex, interconnected, and systemic.

Undoubtedly, because psychological phenomena have traditionally been considered unobservable, there is reason to believe that objectivity may be impossible to achieve. Further, in psychology, mechanical objectivity works for non-complex matters such as the Weber-Fechner law, but more complex psychological theories apparently lack objectivity. This could be one reason why psychology as a discipline suffers most from the replication crisis, as Freese and Peterson (2018) note. The troubleshooting process is intense, but also urgent. The more latent constructs, statistical sophistication, and implicit probability calculations entered the methodological logic of psychological science, and the more theoretical (Fiedler et al., 2021), metrological (Uher, 2018, 2020, 2021), contextual (Borgstede and Scholz, 2021), and epistemological (Meehl, 2009; Hanfstingl, 2019) considerations were ignored, the greater the problem became (e.g., Open Science Collaboration, 2015).

CONSIDERING "PERSPECTIVE"

The term "perspective" has a tradition in psychological sciences, but less on an epistemological level than from a cognitive-developmental psychological perspective. The APA (see text footnote 1) defines perspective as "(1) the ability to view objects, events, and ideas in realistic proportions and relationships; (2) the ability to interpret relative position, size, and distance of objects in a plane surface as if they were three-dimensional; (3) the capacity of an individual to take into account and potentially understand the perceptions, attitudes, or behaviors of themselves and other individuals; and (4) a particular way of looking at events or situations: a stance or philosophical position."

In recent contributions to the philosophy and sociology of science, regarding the objectivity problem, the idea of perspective is discussed epistemologically. For example, Susen (2015) describes the opposition of "perspective versus truth" as an essential criterion of the so-called "postmodern turn" in the social sciences. He argues that perspective could replace the binary concept of objectively true and objectively false in science. Additionally, from a postmodern, feminist, and standpoint tradition, Harding (1995, 2015) suggests using the term "strong objectivity," which means considering traditional scientific objectivity and the perspective of the scientist who achieves a scientific result. Similarly, Daston (1992) speaks of a perspectival objectivity. Tannoch-Bland (1997) claims moral objectivity since morality is often claimed by authorities, depending on historical contexts. These considerations are in line with the ideas by Daston and Galison (2010),Gaukroger (2012), and Tomasello (2020). Epistemologically, there is an agreement that we must not abandon the idea of objectivity, but we have to enrich the original idea with perspective-taking. However, we require a formalized solution, which can be realized on a methodological, empirical, and statistical level.

COLLECTIVE OBJECTIVITY IN A STATISTICAL UNDERSTANDING

For Freese and Peterson (2018), a collective level of objectivity is the only way to escape individual perspectives and subjectivity, and they suggest using meta-analyses to address this issue. The authors call this approach collective or statistical objectivity, seeing meta-analyses as the apex of objectivity (Freese and Peterson, 2018). At first glance, meta-analyses meet the criteria of combining different single studies, therefore different perspectives. It is no coincidence that they are hyped in the current scientific milieu (e.g., Iliescu et al., 2022). Freese and Peterson (2018) argue that single study results often are influenced by scientists' "scientific selves," which, in turn, are affected by different interests, such as emotional or economic. From a cognitive angle, Hanfstingl (2019) mentions the scientific-selves-biases in the work with latent constructs, emphasizing that these problems are grounded to a certain degree on our cognitive automatisms. Meta-analyses, unlike single studies, can reveal statistical effects that would otherwise go undetected.

However, researchers agree that meta-analyses only provide objective knowledge when they are informed by modern quality criteria, such as standardized reporting guidelines and free access to all data and analysis syntaxes, which is often not the case for meta-analyses (Lakens et al., 2016; Polanin et al., 2020). Glass (2015) acknowledges that the initial phase of meta-analyses was characterized by arbitrary decisions and a lack of quality criteria. Still, the provided information is not sufficient to maintain reproducibility and, therefore, higher objectivity (Maassen et al., 2020). Several authors propose clear criteria of reproducibility for future science (e.g., Gurevitch et al., 2018) without considering different perspectives to reach objectivity. Munafò et al. (2017) point to perspective-taking by mentioning team science and the advantages of collaboration, but not systematically and in-depth. However, there are several proposed solutions that have already been published.

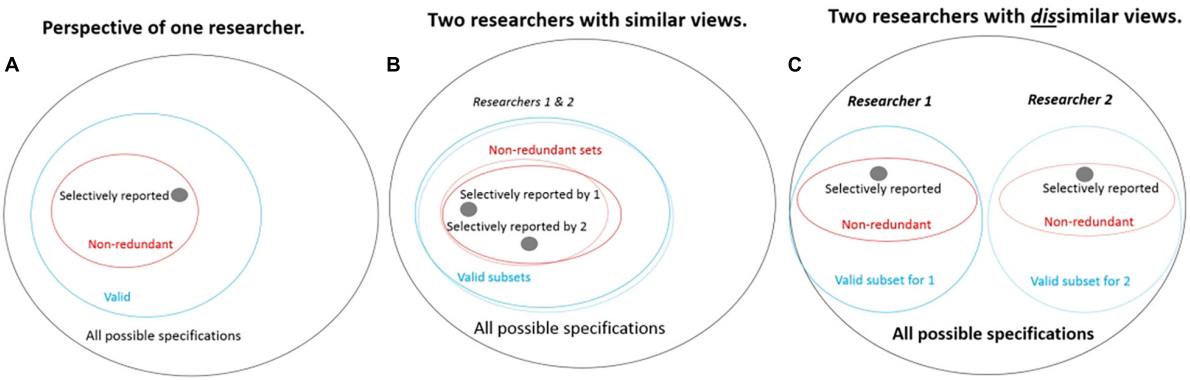


FIGURE 1 | Sets of possible specifications as perceived by researchers. **(A)** The set of specifications reported in an article are a small subset of those the researcher would consider valid to report. **(B)** Different researchers may have similar views on the set of valid specifications but report quite different subsets of them. **(C)** Different researchers may also disagree on the set of specifications they consider valid (Simonsohn et al., 2020, p. 2, **Figure 1**; reprinted with permission by the first author).

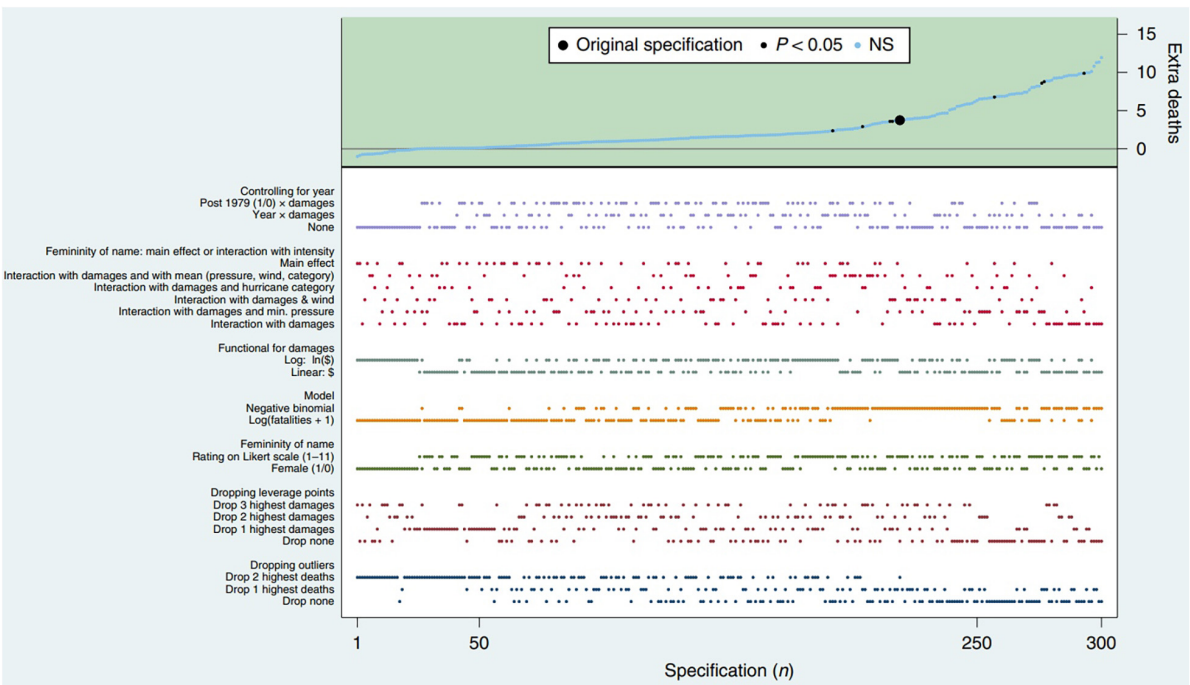


FIGURE 2 | Descriptive specification curve. Each dot in the top panel (green area) depicts the marginal effect, estimated at sample means, of a hurricane having a female rather than male name; the dots vertically aligned below (white area) indicate the analytical decisions behind those estimates. A total of 1,728 specifications were estimated; to facilitate visual inspection, the figure depicts the 50 highest and lowest point estimates and a random subset of 200 additional ones, but the inferential statistics for specification curve analysis include all 1,728 specifications. NS, not significant (Simonsohn et al., 2020, p. 3, **Figure 2**; reprinted with permission by the first author).

Voracek et al. (2019) point to approaches that are able to illustrate researchers' degrees of freedom systematically. For this, they combine solutions developed by Olkin et al. (2012),Steegen et al. (2016), and Simonsohn et al. (2020). Simonsohn et al. (2020) developed their approach for single studies and called it the Specification Curve Analysis, which aims to specify all reasonable and arguable decisions and specifications to answer a research question. The authors also incorporate the problem of different scientists' perspectives (**Figure 1**) and aim to systematically depict them on a "specification curve." These specifications have to (1) sensibly test the research question, (2) be expected to be statistically valid, and (3) not be redundant with other specifications in the set (Simonsohn et al., 2020). The specification curve (**Figure 2**) describes the estimated effect sizes across all specifications, organized around a dashboard chart showing the operationalizations behind each result. Thus, an estimation of the factors influencing the results (decisions on theoretical and methodological approaches, interpretation habits, i.e., scientific selves), can be illustrated in a structured and comprehensible way on the basis of many single studies. Steegen et al. (2016) provide the idea of a multiverse analysis as a similar approach using additional plot alternatives.

Because Steegen et al.'s (2016) and Simonsohn et al.'s (2020) approaches only focus on the single-study-level, Voracek et al. (2019) combine specification curves with an all-subsets combinatorial meta-analysis approach by Olkin et al. (2012). In other words, specification curves combined with combinatorial meta-analyses lead to a systematic overview of possible outcomes resulting from various decisions made by scientists at a meta-meta level. However, Voracek et al. (2019) mention two main problems in their approach. First, the analyses quickly become unfeasible due to the many possible combinations, which could be met with a specific bootstrap strategy they suggest in their manuscript. Second, the approach is still not free of subjective considerations, as "factors need to be tailor-made each time anew, informed by specific debates in the primary literature or by prior related meta-analyses" (Voracek et al., 2019, p. 78). Although researchers get a completely new meta-level of knowledge with this approach family, it seems that the last decision level stays human-biased and perspective-dependent. Nonetheless, this analysis strategy allows many human biases to be made explicitly visible at an individual studies level and meta-analyses in an unprecedented, systematic way. This has been shown by two recent applications of these approaches, both on a meta-meta-level: Dürlinger and Pietschnig's (2022) investigation of the association between intelligence and religiosity, and Vilsmeier et al.'s (2021) analysis of the stability of birth order effects.

CONCLUSION

No scientific method ensures objectivity automatically, and mechanical objectivity is hard to meet for many scientific results. There is a high agreement that objectivity has to be redefined formally. Most authors working on the objectivity problem suggest including perspectives into the concept of objectivity. For example, Harding (2015) says that objectivity is weak as long as we do not consider perspectives; Tannoch-Bland (1997) focuses on moral objectivity when considering authorities' perspectives, both from a feminist context; Susen (2015) contrasts perspective with truth within a postmodern turn, and authors who offer statistical and computational solutions include researchers' degrees of freedom, that is, their perspectives (Olkin et al., 2012; Steegen et al., 2016; Freese and Peterson, 2018; Voracek et al., 2019; Simonsohn et al., 2020). There is also an agreement that, ultimately, influences of human biases remain also on a meta-meta level.

However, is it possible to avoid the frequently-mentioned postmodern arbitraries? I would say, yes, more than that. If we systematically consider objectivity, including diverse perspectives, the validity of science grows instead of shrinks. The more result-biasing specifications and perspectives are known, the more a phenomenon of interest can be disentangled from them. This assumption is supported by recent developments in (psychological) science, in which a major goal is to conduct research in the vein of, for example, an open science policy that can be applied at both single-study and meta-study levels. Many rules were brought together by the open science movement, like ensuring transparency and reproducibility of research processes, preregistrations of studies, or open access to raw data, analysis strategies, syntaxes, and manuscripts. Several older ideas are consistent with systematically accounting for different contextual influences when, for example, randomization tests are used in smaller data sets. Dugard et al. (2012) integrate this idea at the planning stage of a research design, which already implies a prospective validation approach and an orientation toward preregistration, respectively. There is an agreement that it is barely possible to avoid the degrees of freedom when deciding how to frame a study or meta-study. However, as one of the reviewers of this manuscript mentioned, researchers' profound subject matter knowledge helps to use these degrees of freedom in the interest of scientific progress, which goes in line with the argumentation of Hanfstingl (2019).

Open access to research at all its stages opens up the possibility of organizing research programs in such a way that meta-analyses and meta-meta-analyses can be conducted not only backward but forward and prospectively as a regular and thus well-prepared part of objectification and validation processes. Although open access is not necessarily a prerequisite for this consideration, it does bundle together ideas for increasing the objectivity and validity of scientific results. Initiatives, such as big team programs, foster such research strategies and are growing in different fields, be it in general medical science (Steer et al., 2017), psychological science (Forscher et al., 2020), or in a more specific manner, like addiction science (e.g., Pennington et al., 2022). As mentioned above, the troubleshooting process is urgent, but also intense.

AUTHOR CONTRIBUTIONS

BH was the sole author of the manuscript and included the conception and organization of the ideas in the manuscript, recherche of the references, revised, and approved the submitted version.

FUNDING

The publication of the manuscript is funded by the University of Klagenfurt and the Institute of Instructional and School Development.

ACKNOWLEDGMENTS

The author would like to show her gratitude to the editor and reviewers for their insights and constructive feedback. Their provided expertise has improved the quality of the manuscript.

REFERENCES

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F. J., Boker, S. M., Ceulemans, E., and Chen, M. (2020). Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. *J. Psychosom. Res.* 137:110211. doi: 10.1016/j.jpsychores.2020.110211

Borgstede, M., and Scholz, M. (2021). Quantitative and Qualitative Approaches to Generalization and Replication-A Representationalist View. *Front. Psychol.* 12:605191. doi: 10.3389/fpsyg.2021.605191

Daston, L. (1992). Objectivity and the Escape from Perspective. *Soc. Stud. Sci.* 22, 597–618. doi: 10.1177/030631292022004002

Daston, L., and Galison, P. (2010). *Objectivity (First Paperback Edition, Fourth Paperback Printing). Zone Books*. Cambridge: MIT Press.

Dugard, P., File, P., and Todman, J. (2012). *Single-Case And Small-N Experimental Designs*. Oxfordshire: Routledge, doi: 10.4324/9780203180938

Dürlinger, F., and Pietschnig, J. (2022). Meta-analyzing intelligence and religiosity associations: evidence from the multiverse. *PLoS One* 17:e0262699. doi: 10.1371/journal.pone.0262699

Fiedler, K., McCaughey, L., and Prager, J. (2021). Quo Vadis, Methodology? The Key Role of Manipulation Checks for Validity Control and Quality of Science. *Perspect. Psychol. Sci.* 16, 816–826. doi: 10.1177/1745691620970602

Forscher, P. S., Wagenmakers, E., Coles, N. A., Silan, M. A., Dutra, N. B., Basnight-Brown, D., et al. (2020). The Benefits, Barriers, and Risks of Big Team Science. *PsyArXiv* [preprint] doi: 10.31234/osf.io/2mdxh

Freese, J., and Peterson, D. (2018). The Emergence of Statistical Objectivity: changing Ideas of Epistemic Vice and Virtue in Science. *Sociol. Theory* 36, 289–313. doi: 10.1177/0735275118794987

Gaukroger, S. (2012). *Objectivity: a Very Short Introduction. Very Short Introductions*. Oxford: Oxford University Press.

Glass, G. V. (2015). Meta-analysis at middle age: a personal history. *Res. Synth. Methods* 6, 221–231. doi: 10.1002/jrsm.1133

Gurevitch, J., Koricheva, J., Nakagawa, S., and Stewart, G. (2018). Meta-analysis and the science of research synthesis. *Nature* 555, 175–182. doi: 10.1038/nature25753

Hanfstingl, B. (2019). Should We Say Goodbye to Latent Constructs to Overcome Replication Crisis or Should We Take Into Account Epistemological Considerations? *Front. Psychol.* 10:1949. doi: 10.3389/fpsyg.2019.01949

Harding, S. (1995). Strong objectivity? A response to the new objectivity question. *Synthese* 104, 331–349. doi: 10.1007/BF01064504

Harding, S. G. (2015). *Objectivity And Diversity: Another Logic Of Scientific Research*. Chicago: The University of Chicago Press.

Iliescu, D., Rusu, A., Greiff, S., Fokkema, M., and Scherer, R. (2022). Why We Need Systematic Reviews and Meta-Analyses in the Testing and Assessment Literature. *Eur. J. Psychol. Assess.* 38, 73–77. doi: 10.1027/1015-5759/a000705

Lakens, D., Hilgard, J., and Staaks, J. (2016). On the reproducibility of meta-analyses: six practical recommendations. *BMC Psychol.* 4:24. doi: 10.1186/s40359-016-0126-3

Maassen, E., van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A., and Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLoS One* 15:e0233107. doi: 10.1371/journal.pone.0233107

Meehl, P. E. (2009). "The problem is epistemology, not statistics: replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions," in *Multivariate Applications Book series. What if there were no Significance tests?*, eds L. L. Harlow, S. A. Mulaik, and J. H. Steiger (Hove: Psychology Press), 393–452.

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., and Du Sert, N. P. (2017). A manifesto for reproducible science. *Nat. Hum. Behav.* 1:0021. doi: 10.1038/s41562-016-0021

Olkin, I., Dahabreh, I. J., and Trikalinos, T. A. (2012). Gosh - a graphical display of study heterogeneity. *Res. Synth. Methods* 3, 214–223. doi: 10.1002/jrsm.1053

Open Science Collaboration (2015). Psychology. Estimating the reproducibility of psychological science. *Science* 349, aac4716. doi: 10.1126/science.aac4716

Pennington, C. R., Jones, A. J., Tzavella, L., Chambers, C. D., and Button, K. S. (2022). Beyond online participant crowdsourcing: the benefits and opportunities of big team addiction science. *Exp. Clin. Psychopharmacol. Adv.* doi: 10.1037/pha0000541 [Epub ahead of print].

Polanin, J. R., Hennessy, E. A., and Tsuji, S. (2020). Transparency and Reproducibility of Meta-Analyses in Psychology: a Meta-Review. *Perspect. Psychol. Sci.* 15, 1026–1041. doi: 10.1177/1745691620906416

Popper, K. R. (1972). *Objective Knowledge: An Evolutionary Approach*. Oxford: Clarendon Press.

Reiss, J., and Sprenger, J. (2020). *Scientific Objectivity. The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/archives/win2020/entries/scientific-objectivity/ (accessed March 07, 2022).

Rosenthal, R., and Rosnow, R. L. (2009). *Artifacts in Behavioral Research: Robert Rosenthal and Ralph l. Rosnow's Classic Books*. Oxford: Oxford University Press, doi: 10.1093/acprof:oso/9780195385540.001.0001

Schweinsberg, M., Feldman, M., Staub, N., van den Akker, O. R., van Aert, R. C., and van Assen, M. A. (2021). Same data, different conclusions: radical dispersion in empirical results when independent analysts operationalize and test the same hypothesis. *Organ. Behav. Hum. Decis. Process.* 165, 228–249. doi: 10.1016/j.obhdp.2021.02.003

Silberzahn, R., Uhlmann, E. L., Martin, D. P., Anselmi, P., Aust, F., and Awtrey, E. (2018). Many Analysts, One Data Set: making Transparent How Variations in Analytic Choices Affect Results. *Adv. Meth. Pract. Psychol. Sci.* 1, 337–356. doi: 10.1177/2515245917747646

Simonsohn, U., Simmons, J. P., and Nelson, L. D. (2020). Specification curve analysis. *Nat. Hum. Behav.* 4, 1208–1214. doi: 10.1038/s41562-020-0912-z

Steegen, S., Tuerlinckx, F., Gelman, A., and Vanpaemel, W. (2016). Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci.* 11, 702–712. doi: 10.1177/1745691616658637

Steer, C. J., Jackson, P. R., Hornbeak, H., McKay, C. K., Sriramarao, P., and Murtaugh, M. P. (2017). Team science and the physician-scientist in the age of grand health challenges. *Ann. N. Y. Acad. Sci.* 1404, 3–16. doi: 10.1111/nyas.13498

Susen, S. (2015). *The Postmodern Turn In The Social Sciences*. London: Palgrave Macmillan.

Tannoch-Bland, J. (1997). From Aperspectival Objectivity to Strong Objectivity: the Quest for Moral Objectivity. *Hypatia* 12, 155–178. doi: 10.1111/j.1527-2001.1997.tb00176.x

Tomasello, M. (2020). The role of roles in uniquely human cognition and sociality. *J. Theory Soc. Behav.* 50, 2–19. doi: 10.1111/jtsb.12223

Uher, J. (2018). Quantitative Data From Rating Scales: an Epistemological and Methodological Enquiry. *Front. Psychol.* 9:585. doi: 10.3389/fpsyg.2018.02599

Uher, J. (2020). Measurement in metrology, psychology and social sciences: data generation traceability and numerical traceability as basic methodological principles applicable across sciences. *Qual. Quant.* 54, 975–1004. doi: 10.1007/s11135-020-00970-2

Uher, J. (2021). Psychometrics is not measurement: unraveling a fundamental misconception in quantitative psychology and the complex network of

its underlying fallacies. *J. Theor. Philos. Psychol.* 41, 58–84. doi: 10.1037/teo0000176

Vilsmeier, J. K., Kossmeier, M., Voracek, M., and Tran, U. S. (2021). The Fraternal Birth-Order Effect as Statistical Artefact: convergent Evidence from Probability Calculus, Simulated Data, and Multiverse Meta-Analysis. *PsyArXiv* [preprint] doi: 10.31234/osf.io/e4j6a

Voracek, M., Kossmeier, M., and Tran, U. S. (2019). Which Data to Meta-Analyze, and How? *Z. Für Psychol.* 227, 64–82. doi: 10.1027/2151-2604/a000357

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychology

TYPE Opinion
PUBLISHED 13 October 2022
DOI 10.3389/fpsyg.2022.986767

OPEN ACCESS

EDITED BY
Chris Evans,
Universidad de Las Américas (UDLA), Ecuador

REVIEWED BY
Jörg Henseler,
University of Twente, Netherlands

*CORRESPONDENCE
Peter Adriaan Edelsbrunner
peter.edelsbrunner@ifv.gess.ethz.ch

SPECIALTY SECTION
This article was submitted to Quantitative Psychology and Measurement, a section of the journal Frontiers in Psychology

RECEIVED 05 July 2022
ACCEPTED 28 September 2022
PUBLISHED 13 October 2022

CITATION
Edelsbrunner PA (2022) A model and its fit lie in the eye of the beholder: Long live the sum score. *Front. Psychol.* 13:986767. doi: 10.3389/fpsyg.2022.986767

A model and its fit lie in the eye of the beholder: Long live the sum score

Peter Adriaan Edelsbrunner*

Department of Humanities, Social and Political Sciences, ETH Zurich, Zurich, Switzerland

KEYWORDS

latent variable, sum score, formative measurement, model equivalence, socio-constructivism

Introduction

A common way to build scores for statistical analysis from psychological or educational scales (e.g., knowledge or intelligence tests, attitude or motivation questionnaires,) is to sum up participants' scores across all items. For example, in a questionnaire assessing knowledge about a specific topic, participants might have to answer different questions probing their content knowledge. The correct answers on all items are then summed up to build a score for analysis that is meant to represent participants' knowledge (Edelsbrunner et al., 2018). Similarly, on an intelligence test, correctly solved items might be summed to yield an overall score that is transformed into IQ estimates (Raven et al., 1962). On a questionnaire measuring need for cognition, participants might indicate their agreement with different self-descriptions on a Likert scale and their agreement is summed up (or a mean is built, which is equivalent for the arguments brought up here) across all items (Beißert et al., 2014).

Reasons to follow this practice and using sum (or mean) scores include the statistical and conceptual simplicity of building such a score: building a sum score does not require setting up an elaborate statistical model, and it might appear easily defensible to just follow this common practice without further notice.

Recent literature argues that building a sum score is, statistically and conceptually, not as innocent as it seems (Kuhfeld and Soland, 2020; McNeish and Wolf, 2020). Specifically, McNeish and Wolf (2020) argued that a sum score, although not involving any explicit statistical model in its computation, implicitly assumes a very specific and very stringent statistical model. The authors argue that whenever researchers use a sum score, they implicitly assume that a variant of factor analysis fits their data that imposes equal factor loadings and error variances across all items (the *parallel* factor model). They note that although seeming like a simple arithmetic operation, sum scoring actually *is* a simple transformation of this model. They further argue that actual factor-analytic methods have been shown to provide more accurate estimates. Consequently, they infer that using sum scores obliges researchers to engage with and test the model constraints implied by sum scores, because as in physical or social sciences, no conclusions would be endorsed without evidence (McNeish and Wolf, 2020).

I provide three counter-arguments against these views, arguing that (1) factor analysis makes meta-theoretical assumptions that are extremely stringent and seldom justified although they might not fit many constructs, (2) sum scores do not imply a factor model, because there are many other known models that imply the same data patterns as factor analysis, and (3) there is an infinite number of further models that imply the same data. Overall, this will bring me to the point that sum scores, as well as any other way to build scores, only imply the models that are theoretically defensible, not empirically.

Meta-theoretical assumptions of factor analysis

With meta-theoretical assumptions, I describe theoretical assumptions implicit to the defining parameter structure of a class of statistical model. Factor analysis implies the meta-theoretical assumption that the latent construct that we aim at measuring is reflective (MacKenzie et al., 2005). This means that whatever the latent variable (i.e., the factor representing the latent construct; for a discussion see Maraun and Halpin, 2008) represents causally influences its indicator variables (i.e., items), and that all variance in the indicators that they do not share is measurement error. It is well-known that these assumptions do not appear appropriate for many constructs in education and psychology, theoretically (van der Maas et al., 2006) and empirically (van der Maas and Kan, 2016; van Bork et al., 2021).

Whether these assumptions are theoretically defensible can be answered by two simple questions. First, does it make sense to assume that the latent construct influences its indicators causally (e.g., if I am more motivated, this will affect me to agree more strongly with all items on a motivation-questionnaire)? If other causal pathways appear more reasonable, than the reflectivity-assumption might not be a useful representation of data. For example, instead of the construct influencing its indicators, for some constructs it might be more reasonable to assume the other causal pathway, that it, the construct being influenced by its indicators. The indicators educational background and salary raise my socio-economic status, instead of my socio-economic status raising my education and salary (Schuberth, in press). Another possibility is that there is no unifying latent construct; instead, the indicators might directly influence one another. For example, being sleepless might cause anhedonia directly, rather than both being explained by a latent construct of depression (van de Leemput et al., 2014). If any of these alternative assumptions about the relation between a construct and its indicators appears more reasonable, then factor analysis might not be a very informative model of the data-generating mechanism underlying a sum score.

A second question that might be posed to evaluate whether a factor analytic model is an appropriate representation of a construct is whether the items that were developed to capture the construct are replaceable with one another. In factor analysis, exchanging or using only a subset of indicators is supposed not to alter the meaning of the construct (White et al., 2022). If this is not the case, for example because sleeplessness and anhedonia each provide important information about depression beyond each other, than a model representing one of the outlined alternative kinds of constructs might be more appropriate than factor analysis.

Alternative models producing the same data

The assumption that a sum score is "a simple linear transformation of a heavily constrained parallel factor model" (McNeish and Wolf, 2020) is questionable, given that a row of other statistical models have been shown to imply very similar data patterns as factor analysis. From a logical perspective, this is a converse error ("affirming the consequent"; see e.g., Martinsen, 2022): given that a constrained version of factor analysis implies a sum score, McNeish and Wolf assume that a sum score must imply factor analysis, overlooking that other models could also imply sum scores. Specifically, it can be shown mathematically that data patterns implied by factor analysis are also implied by multiple other kinds of models (e.g., Schuberth, in press). For example, a latent class analysis, modeling two classes of individuals through a categorical rather than a continuous latent variable, will generally imply the same means and variance-covariance matrix as a unidimensional factor analysis applied to the same data (Molenaar and von Eye, 1994). In addition, it has been shown that psychometric network models can produce data that are in accordance with factor analysis (van der Maas et al., 2006). Another kind of model implying equivalent data is a composite, which conceptualizes a formative rather than a reflective construct (Schuberth, in press). It stands to debate why exactly factor analysis should be applied to data to see whether it fits data and to extract factor sores, given that these alternative models, particularly if they are in better accordance with theoretical assumptions, might better capture the data-generating process.

An infinite number of other models will fit the same data

Beyond models that are already known to mankind, we can be quite sure that many further psychometric models will be developed in the time to come. It has been shown that in principle, for any model an infinite number of alternative models exist that can fit an observed variance-covariance matrix equally well (e.g., Raykov and Marcoulides, 2001). I therefore suggest not considering any model that might be the prevalent "best practice" at one point in time as *the* data-generating

model behind a sum score, or any other kinds of scores. As long as methodological research continues, it will develop many new informative models that will provide reasonable accounts of data just as well as, or even better than, factor analysis. This is especially noteworthy given that factor analysis was established more than 100 years ago (Spearman, 1904), and by now we already have a number of alternative models to choose from. Perhaps, factor analysis is commonly assumed to underlie data mostly because it has been around for so long and is a comparably well-developed approach.

Alternative justifications of sum scores

Do I believe that these arguments free researchers from any justifications for their uses of sum scores? No. To the contrary, I would like to bring up alternative justifications. First, a sum score can be built if researchers have conceptualized the construct that they intend to represent in the score such that all indicators represent approximately equal shares of the construct. This might for example be the case if researchers aim at building an index of a construct of different skills (Van der Maas et al., 2014) or beliefs (e.g., Merk and Rosman, 2019; Schiefer et al., 2022). In this case, however, it should not be overlooked that equal weighting of indicators in an index should also be based on justification. One such justification might be that different aspects of a construct have been defined as equally important components of a theoretical model, or that they are assumed to play similarly important roles in determining an educational or psychological outcome. In such cases, a sum score provides good construct representation, that is, match between the meta-theoretical assumptions of a statistical model and the theoretical construct that it is meant to represent.

A common case probably is that researchers think about the different aspects that the construct they are intending to measure consists of, and then they develop indicators such as items in a balanced manner, such that each of these aspects is represented by the same number of items. It stands to debate why in such cases, any item(s) should be weighted more strongly than others.

Another reason to use sum scores is the aim to use scores that are as comparable across studies as possible in their constituents. If factor analysis is applied to data, factor loadings are usually estimated in a data-driven manner, such that items are put into the score with weights that have little theoretical justification. This will usually induce variation in scores that are extracted based on factor analysis across studies: if in one study, an indicator has received a strong factor weight but in another study, its weight is lower, then it is difficult to compare the meaning of these derived scores across these studies conceptually (Widaman and Revelle, in press). This might also be described as a classical issue of a bias-variance tradeoff (Yarkoni and Westfall, 2017). If scores are built in an equal manner across studies, for example by always using sum scores, then they might be biased if factor analysis might have been an unbiased model of the data-generating process. At the same time, the variance in the composition of the scores will be lower if they are always built alike. Consider however, perhaps speaking against this argument, the point that if a factor analytic model fits data well, the meaning of the latent variable does not change when indicators are replaced with one another. Consequently, also if factor loadings differ between studies, this might not change the meaning of the latent variable. This assumption might be valid as long as at least some indicators have consistent loadings across studies, implying partial measurement invariance (although measurement invariance does not imply equality of latent variables; Maraun and Heene, 2016).

Even if researchers use sum scores with theoretical but without statistical justification, this might be defensible. Educational and psychological science generally follow the maxim of an empirical science grounded in modernism (Holtz, 2020); data should inform theories. This does not, however, have to mean that all uses of data have to be empirically justified in all cases. Education and psychology are not just empirical but also socio-constructivist sciences (Guyon and Nôus, 2021). Within such a science, instead of justifying use of data exclusively empirically, researchers should be free to justify whether their use of data should be based on empirical fit, or on theoretical fit, or on both.

Finally, if researchers decide to base their use of data on empirical fit, some approaches have been developed to differentiate between data patterns that the discussed kinds of models typically produce. For example, throughout the last decades, various statistical approaches for distinguishing between factor analytic and latent class models have been introduced (De Boeck et al., 2005), as well as for distinguishing between network models and latent factors (van Bork et al., 2021).

Take-home message

Overall, I agree with McNeish and Wolf (2020) in asking researchers for justifications for their uses of scores. I am, however, going a step back (or further, depending on the eye of the beholder) and asking researchers to first justify which of their procedures they want to justify theoretically, and which empirically, and why. Theoretical justification can be achieved through conceptualization and definition. If a construct is defined in a way such that the building of a sum score maps on this definition well (Lundberg et al., 2021), then its use is appropriate.

What will happen if in a 100 years, factor analysis is not used anymore but has been superseded by a new class of models that takes very different conceptual and statistical perspectives?

And, if some variant of these new models also implies a sum score, does this then mean that sum scores will represent that new future model? Probably not. Sum scores represent only the model that has been used to create them for good theoretical reasons.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

Open access funding provided by ETH Zurich.

Acknowledgments

I would like to thank Daniel McNeish, Clemens Lechner, and Jonas Haslbeck for fruitful discussions, as well as the reviewer for bringing up valuable points.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Beißert, H., Köhler, M., Rempel, M., and Beierlein, C. (2014). *Eine Deutschsprachige Kurzskala zur Messung des Konstrukts Need for Cognition: Die Need for Cognition Kurzskala (NFC-K). GESIS-Working Papers 2014/32*. Available online at: http://nbn-resolving.de/urn:nbn:de:0168-ssoar-403157

De Boeck, P., Wilson, M., and Acton, G. S. (2005). A conceptual and psychometric framework for distinguishing categories and dimensions. *Psychol. Rev.* 112, 129. doi: 10.1037/0033-295X.112.1.129

Edelsbrunner, P. A., Schalk, L., Schumacher, R., and Stern, E. (2018). Variable control and conceptual change: a large-scale quantitative study in elementary school. *Learn. Individ. Differ.* 66, 38–53. doi: 10.1016/j.lindif.2018.02.003

Guyon, H., and Nôus, C. (2021). Quantification in experimental psychology and pragmatic epistemology: tension between the scientific imperative and the social imperative. *Front. Psychol.* 11, 603617. doi: 10.3389/fpsyg.2020.603617

Holtz, P. (2020). Does postmodernism really entail a disregard for the truth? Similarities and differences in postmodern and critical rationalist conceptualizations of truth, progress, and empirical research methods. *Front. Psychol.* 11, 545959.

Kuhfeld, M., and Soland, J. (2020). Avoiding bias from sum scores in growth estimates: An examination of IRT-based approaches to scoring longitudinal survey responses. *Psychol. Methods* 27, 234. doi: 10.1037/met0000367

Lundberg, I., Johnson, R., and Stewart, B. M. (2021). What is your estimand? Defining the target quantity connects statistical evidence to theory. *Am. Sociol. Rev.* 86, 532–565. doi: 10.1177/0003122421100418 7

MacKenzie, S. B., Podsakoff, P. M., and Jarvis, C. B. (2005). The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *J. Appl. Psychol.* 90, 710–730. doi: 10.1037/0021-9010.90.4.710

Maraun, M., and Heene, M. (2016). A note on the implications of factorial invariance for common factor variable equivalence. *Commun. Stat. Theor. Methods* 45, 4181–4193. doi: 10.1080/03610926.2014.917186

Maraun, M. D., and Halpin, P. F. (2008). Manifest and latent variates. *Measurement: Interdisciplinary Research and Perspectives* 6, 113–117. doi: 10.1080/15366360802035596

Martinsen, T. (2022). Developing critical thinking military officers. *Mathematica Militaris* 25, 2. Available online at: https://digitalcommons.usmalibrary.org/mathematica_militaris/vol25/iss1/2

McNeish, D., and Wolf, M. G. (2020). Thinking twice about sum scores. *Behav. Res. Methods* 52, 2287–2305. doi: 10.3758/s13428-020-01398-0

Merk, S., and Rosman, T. (2019). Smart but evil? student-teachers' perception of educational researchers' epistemic trustworthiness. *AERA Open* 5, 1–18. doi: 10.1177/2332858419868158

Molenaar, P. C. M., and von Eye, A. (1994). "On the arbitrary nature of latent variables," in *Latent Variables Analysis: Applications for Developmental Research* (Thousand Oaks, CA: Sage Publications, Inc.), 226–242.

Raven, J. C., Raven, J. C., and John Hugh Court (1962). *Advanced Progressive Matrices*. London: HK Lewis.

Raykov, T., and Marcoulides, G. A. (2001). Can there be infinitely many models equivalent to a given covariance structure model? *Structural Equation Modeling*. 8, 142–149. doi: 10.1207/S15328007SEM0801_8

Schiefer, J., Edelsbrunner, P. A., Bernholt, A., Kampa, N., and Nehring, A. (2022). Epistemic beliefs in science—a systematic integration of evidence from multiple studies. *Educ. Psychol. Rev.* 34, 1–35. doi: 10.1007/s10648-022-09661-w

Schuberth, F. (in press). The Henseler-Ogasawara specification of composites in structural equation modeling: a tutorial. *Psychol. Methods*. doi: 10.1037/met0000432

Spearman, C. (1904). "General intelligence," objectively determined and measured. *Am. J. Psychol.* 15, 201–292. doi: 10.2307/1412107

van Bork, R., Rhemtulla, M., Waldorp, L. J., Kruis, J., Rezvanifar, S., and Borsboom, D. (2021). Latent variable models and networks: statistical equivalence and testability. *Multivar. Behav. Res.* 56, 175–198. doi: 10.1080/00273171.2019.1672515

van de Leemput, I. A., Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., et al. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences* 111, 87–92. doi: 10.1073/pnas.1321114110

Van der Maas, H. L., Kan, K. J., and Borsboom, D. (2014). Intelligence is what the intelligence test measures. Seriously. *J. Intelligence* 2, 12–15. doi: 10.3390/jintelligence2010012

van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., and Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: the positive manifold of intelligence by mutualism. *Psychol. Rev.* 113, 842–861. doi: 10.1037/0033-295X.113.4.842

van der Maas, H. L. J., and Kan, K. J. (2016). Comment on "Residual group-level factor associations: possibly negative implications for the mutualism theory of general intelligence" by Gilles E. Gignac (2016). *Intelligence 57*, 81–83. doi: 10.1016/j.intell.2016.03.008

White, M., Edelsbrunner, P. A., and Thurn, C. (2022). The concept(ualization) implies the statistical model: implications for measuring teaching quality. *Preprint, available from PsyArXiv [preprint]*. doi: 10.31219/osf.io/zksbt

Widaman, K. F., and Revelle, W. (in press). Thinking thrice about sum scores, and then some more about measurement and analysis. *Behav. Res. Methods* 1–19. doi: 10.3758/s13428-022-01849-w. [Epub ahead of print].

Yarkoni, T., and Westfall, J. (2017). Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect. Psychol. Sci.* 12, 1100–1122. doi: 10.1177/1745691617693393

frontiers | Frontiers in Psychology

TYPE Methods
PUBLISHED 28 December 2022
DOI 10.3389/fpsyg.2022.1009893

OPEN ACCESS

EDITED BY
Tomer Fekete,
Ben-Gurion
University of the Negev, Israel

REVIEWED BY
M. K. Samal,
Bhabha Atomic Research Centre (BARC),
India
David Witherington,
University of New Mexico, United States

*CORRESPONDENCE
Jana Uher
mail@janauher.com

SPECIALTY SECTION
This article was submitted to
Theoretical and Philosophical Psychology,
a section of the journal
Frontiers in Psychology

RECEIVED 02 August 2022
ACCEPTED 14 October 2022
PUBLISHED 28 December 2022

CITATION
Uher J (2022) Rating scales institutionalise a network of logical errors and conceptual problems in research practices: A rigorous analysis showing ways to tackle psychology's crises.
Front. Psychol. 13:1009893.
doi: 10.3389/fpsyg.2022.1009893

Rating scales institutionalise a network of logical errors and conceptual problems in research practices: A rigorous analysis showing ways to tackle psychology's crises

Jana Uher[1,2]*

[1]School of Human Sciences, University of Greenwich, London, United Kingdom, [2]London School of Economics, London, United Kingdom

This article explores in-depth the metatheoretical and methodological foundations on which rating scales—by their very conception, design and application—are built and traces their historical origins. It brings together independent lines of critique from different scholars and disciplines to map out the problem landscape, which centres on the failed distinction between psychology's study phenomena (e.g., experiences, everyday constructs) and the means of their exploration (e.g., terms, data, scientific constructs)—psychologists' cardinal error. Rigorous analyses reveal a dense network of 12 complexes of problematic concepts, misconceived assumptions and fallacies that support each other, making it difficult to be identified and recognised by those (unwittingly) relying on them (e.g., various forms of reductionism, logical errors of operationalism, constructification, naïve use of language, quantificationism, statisticism, result-based data generation, misconceived nomotheticism). Through the popularity of rating scales for efficient quantitative data generation, uncritically interpreted as psychological measurement, these problems have become institutionalised in a wide range of research practices and perpetuate psychology's crises (e.g., replication, confidence, validation, generalizability). The article provides an in-depth understanding that is needed to get to the root of these problems, which preclude not just measurement but also the scientific exploration of psychology's study phenomena and thus its development as a science. From each of the 12 problem complexes; specific theoretical concepts, methodologies and methods are derived as well as key directions of development. The analyses—based on three central axioms for transdisciplinary research on individuals, (1) complexity, (2) complementarity and (3) anthropogenicity—highlight that psychologists must (further) develop an explicit metatheory and unambiguous terminology as well as concepts and theories that conceive individuals as living beings, open self-organising systems with complementary phenomena and dynamic interrelations across their multi-layered systemic contexts—thus, theories not simply of elemental properties and structures but of processes, relations, dynamicity, subjectivity, emergence, catalysis and transformation. Philosophical and theoretical foundations of approaches suited for exploring these phenomena must

be developed together with methods of data generation and methods of data analysis that are appropriately adapted to the peculiarities of psychologists' study phenomena (e.g., intra-individual variation, momentariness, contextuality). Psychology can profit greatly from its unique position at the intersection of many other disciplines and can learn from their advancements to develop research practices that are suited to tackle its crises holistically.

KEYWORDS

measurement, quantitative, psychometrics, replication, validity, generalizability, construct, rating scale

Rating 'scales': Promises and challenges

Psychology is in crisis, again and anew. Continued debates about replicability (Regenwetter and Robinson, 2017), validity (Newton and Baird, 2016), generalisability (Yarkoni, 2022), robust results (Nosek et al., 2022), preregistration (Szollosi et al., 2020), measurement theories (Trendler, 2019; Uher, 2021c,d) and measurability (Michell, 1997; Trendler, 2009), amongst others, indicate profound problems still unsolved. Astonishingly, however, the widespread use of rating 'scales' for quantitative investigations of the complex phenomena of behaviour, psyche and society seems largely unchallenged—even by critics of contemporary practices (e.g., Michell, 2013).

Ratings are popular. Their efficiency to produce large numerical data sets about psychological study phenomena is enormous. Millions of individuals can be studied without any direct contact, nowadays facilitated through online platforms and commercial participant samples featuring desired characteristics (e.g., Amazon's Mechanical Turk; Anderson et al., 2018). Well-trained scientific staff, expensive equipment, or technical measuring devices are not needed. With ratings, ordinary laypeople can produce numerical data. This spares costly efforts otherwise needed to bring individuals to the lab, elicit the phenomena of interest in experiments, wait for their occurrence in field observations, or deal with the ethical intricacies involved in recording individuals' everyday activities. Small sample sizes, intense observer training, complicated experimental setups, laborious (e.g., software-based) coding work from multiple observers contrast with the ease of producing large data sets with just some ticks obtainable any time and (almost) any place.

With ratings, it seems, behaviours can be studied even in retrospect (e.g., habitual behaviours in 'personality' ratings). Behaviour[1] researchers, by contrast, must actually see the behaviours and individuals studied—for ratings, this is not needed. Moreover, ratings are used to assess what even the most meticulous recording of physically described and situationally located behavioural acts cannot capture—their appraisal (e.g., normativity, social valence) and interpretation, such as regarding individuals' intentions, beliefs or feelings that can only be inferred or require self-report. All this information is collected in well-structured data sets, straightforwardly applicable to statistical analysis and seemingly comparable across studies, thus facilitating the generalisation of findings. Compare this with the efforts needed to recruit, meet and interview individuals in one–to–one sessions, to transcribe their verbal data, to code and interpret the textual data thus-produced, and all these efforts to study just small samples with limited options for quantification, comparability and generalisability. No wonder rating 'scales' are popular. Indeed, what else could be done with comparable ease and efficiency?

Ease and efficiency—although relevant given limited resources—have never been hallmarks of scientific excellence. Other sciences invested enormous efforts to enable ever more accurate measurement (e.g., 18th century metrologists[2] measured half the globe to determine the universally agreed length of one metre), to make accessible phenomena previously unexplored (e.g., electron microscope), and to continuously refine their methods to capture even minuscule changes in their objects of research (e.g., spike protein mutations in Sars-Cov2-virus variants). But since the advent of rating methods about a century ago (Thurstone, 1928; Likert, 1932), little has changed in their use to generate data (apart from their digital implementation)—much in contrast to the significant advances made in the statistical analysis and modelling of numerical data thus-produced. Still today, statements or questions (items) describing phenomena of interest are presented to laypersons for graded judgement using fixed answer categories indicating staged degrees (e.g., of frequency) that are commonly considered a 'scale' (e.g., 'rarely', 'sometimes', 'often'). To enable their application to a broad range of phenomena, contexts and individuals without specifying any

1 The study of situated and physically described behaviour as done, for example, in human ethology, child research and behavioural ecology. Not to be confused with behavioural economics, sometimes labelled 'behavioural' science, which largely relies on rating 'scales' and surveys.

2 Metrology, the science of measurement, foundational for the physical sciences and engineering.

particular ones, rating 'scales' are commonly broadly worded (Borkenau and Müller, 1991). Colloquial language is used to ensure these 'scales' are self-explanatory for laypersons. To further simplify their task, items comprise only short phrases or single words that describe only a particular aspect of complex phenomena, thus presenting chunks of information that can be managed efficiently—mentally by raters and analytically by researchers. The items' presentation in a predetermined, mixed order is meant to help raters focus on one item at a time without cross-checking their answers for consistency. Indeed, raters are often encouraged to not ponder too long about an item and to indicate the first answer that comes to their mind. Raters need not even formulate their answers themselves but just to tick off the answer categories provided. Raters' task, so it seems, could not be made any easier.

The apparent simplicity, however, masks intricate challenges imposed on raters. First, raters must interpret items and answer categories to identify relevant phenomena to be judged (e.g., specific behaviours) and the kind of grading enquired (e.g., frequency). Colloquially worded items, however, reflect broad semantic fields of meaning, which are inherently context-dependent. Therefore, raters must construe for each rating a specific meaning and consider specific phenomena to be judged (Wagoner and Valsiner, 2005; Rosenbaum and Valsiner, 2011; Lundmann and Villadsen, 2016; Uher and Visalberghi, 2016). To assess their current intensity or typicality for an individual, raters must recall, consider and weigh relevant occurrences across different occasions, contexts and even individuals (e.g., for 'personality' ratings), form an overall judgement and fit it into the answer 'scale' provided (Uher, 2018a, 2021d). But occurrences of behaviour are highly complex on all levels of consideration (e.g., individuals, situations, groups, time; Uher, 2015b), not to mention the many interpretive perspectives one can take for explaining behaviours, such as regarding possibly underlying intentions, goals or feelings. Considering all this on demand and out of context in a longer series of brief, isolated and broadly worded descriptions and without much reflection is quite challenging. No wonder respondents often use mental shortcuts, consider just single pieces of information or rely on semantic similarity, common stereotypes or answer tendencies (Shweder, 1977; Wood et al., 2012; Uher et al., 2013b; Arnulf et al., 2014; Uher, 2018a), leading to countless well-described biases (Tourangeau et al., 2000; Podsakoff et al., 2003).

All this questions the accuracy of rating data for psychological 'measurement' as well as their utility for quantitative research on the phenomena of behaviour, psyche and society.

This article

This article analyses in-depth the metatheoretical and methodological foundations on which rating 'scale' methods—by their very conception, design and application—are built. *Metatheory* concerns the philosophical and theoretical assumptions that are made about the study phenomena's nature and the questions that can be asked about them. *Methodology* concerns the philosophy and theory of the approaches (ways) and methods that are suited to explore these phenomena. *Methods,* in turn, are the specific practices, procedures and techniques that are used to perform the therefore necessary operations (Althusser and Balibar, 1970; Sprung and Sprung, 1984; Kothari, 2004). Methodology and method are often conflated (especially in English-language psychology). This reflects many psychologists' reluctance to elaborate the philosophical and theoretical foundations of their research practices. Rating methods, for example, are well elaborated but their underlying methodology is not.

The first section outlines the philosophical and conceptual frameworks on which the present analyses are based. This prepares the ground to analyse, in the second section, the conceptual foundations of rating 'scales', whereby independent lines of critique from different scholars and disciplines are integrated and complemented with novel ones. The analyses reveal a network of 12 complexes of problematic conceptions and erroneous assumptions that support each other and that are codified in common psychological jargon, making it difficult to be identified and recognised by those (unwittingly) relying on them. Specifically, the conceptual problems entail logical gaps that are masked by ambiguous terms, which invite conflations of their disparate meanings. This necessitates a conceptual back-and-forth switching between disparate elements of research as an intuitive attempt to bridge these gaps. This conceptual switching is similar to that experienced with ambiguous images (reversible figures), which cause multi-stable perception and illusions. But unlike these perceptual illusions and concealed by the ambiguous terminology, this conceptual back-and-forth switching goes largely unnoticed—as does its failure to remedy the logical problems.

Through the widespread and uncritical application of rating 'scales' as methods enabling psychological 'measurement', these problems have become institutionalised in a wide range of research practices, impacting even scientific activities that should be independent of data generation methods (e.g., choice of research questions). Institutionalised problems cannot be remedied with little quick fixes that many may hope for. The aim is therefore to map out the problem landscape to enable an in-depth understanding of the underlying assumptions and concepts that keep the current problematic research practices in place. In-depth understanding is essential to derive meaningful directions for future developments that are needed to tackle psychology's crises holistically and that are outlined in the final section.

The present analyses: Conceptual foundations

To critically analyse the philosophical and theoretical foundations of a research system, the most general assumptions

on which these analyses are based must be explicated. They form the axiomatic basis from which the specific assumptions and concepts that are used in these analyses are derived and on which the resulting conclusions are based (Collingwood, 1940). Such explications are not commonly made in psychology, which is symptomatic for the discipline's neglect of its own philosophical and theoretical foundations. Psychology has been operating for too long on the basis of implicit paradigms that are taken for granted and no longer considered explicitly, thereby relying on too many (meanwhile) hidden assumptions that urgently need reappraisal, critical reflection and even radical change and renewal (Danziger, 1979; Gergen, 2001; Fahrenberg, 2015; Smedslund, 2016; Valsiner, 2017b; Toomela, 2018). Explicating the philosophical and theoretical foundations helps identify where differences in conception and understanding may originate from and highlights problems and inconsistencies in the conceptual foundations of rating 'scales' but also suitable alternatives.

Transdisciplinary Philosophy-of Science Paradigm for Research on Individuals (TPS-Paradigm)

The present analyses are based on the *Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals* (TPS-Paradigm; for introductory overviews,[3] see Uher, 2015c, 2018a, 2021b, pp. 219–222). The TPS-Paradigm is targeted at making explicit the most basic assumptions that different disciplines (e.g., psychology, biology, medicine, social sciences, physical sciences, metrology) make about research on individuals involving phenomena from all domains of life (e.g., abiotic, biotic, psychical, socio-cultural). Their holistic investigation, necessitated by their joint emergence in the single individual, poses challenges because different phenomena require different epistemologies, theories, methodologies and methods, which are based on different and even contradictory basic assumptions. To provide conceptual foundations that are suitable for tackling these challenges, established concepts from various disciplines have been systematically integrated on the basis of their underlying rationales and basic assumptions, and complemented by novel ones, thereby creating philosophical, metatheoretical and methodological frameworks that coherently build upon each other and that transcend disciplinary boundaries (Figure 1). These frameworks help scientists to critically reflect on, discuss and refine their theories and practices and to develop new ones, and are therefore well-suited for the present analyses.

The TPS-Paradigm's philosophical framework with its three basic presuppositions is outlined now. Relevant concepts from the metatheoretical and methodological frameworks are introduced below where needed, including metatheoretical terms and definitions that are used in the TPS-Paradigm to improve clarity and avoid jingle–jangle fallacies.[4]

Three basic presuppositions

The philosophical framework specifies three central presuppositions that function as the TPS-Paradigm's most basic axioms for research on individuals—(1) complexity, (2) complementarity and (3) anthropogenicity (Figure 1).

Axiom 1 – Complexity: Individuals are complex living systems

As living organisms, individuals are conceived as *open (dissipative) systems* that are in constant exchange with their surroundings but able to maintain themselves through *self-organisation*. Living systems are composed of interrelated elements that are nested on different levels of organisation. On each level, they function as wholes from which new properties emerge that are not predictable from their constituents and that can feed back to the constituents from which they *emerge* (retroaction/transaction), leading to *dynamic, non-linear, dialectical and irreversible processes* of development. With increasing levels of organisation, ever more complex systems emerge that are less rule-bound, highly adaptive and historically unique. Complex psychical systems enable human individuals to be self-reflective, thinking and intentional agents who hold inherently subjective views given their own situatedness in their systemic multi-level contexts (von Bertalanffy, 1937; Shweder, 1977; Capra, 1997; Prigogine and Stengers, 1997; Smedslund, 2004; Morin, 2008; Valsiner, 2021).

Axiom 2 – Complementarity: Different approaches can reveal contradictory information about the same object of research

Particular objects of research can be exhaustively understood only by describing two mutually exclusive properties that are categorically different, maximally incompatible with one another, and neither reducible nor transformable into each other, thus requiring different frames of reference, criteria of 'truth' and methods of investigation, and that may therefore be regarded as *complementary* to one another. This principle was applied to the wave–particle dilemma in research on the nature of light and has

3 The TPS-Paradigm has already been applied to integrate, expand and complement established (1) concepts of psyche, behaviour, language and contexts (e.g., Uher, 2013, 2016a,b); (2) concepts and methodologies for taxonomising and comparing individual differences in various kinds of phenomena within and across populations (e.g., Uher et al., 2013a; Uher, 2015b,c,d,e, 2018b), as well as (3) concepts and theories of data generation, quantification and measurement across the sciences (e.g., Uher, 2019, 2020a,b) and in quantitative psychology and psychometrics (e.g., Uher et al., 2013b; Uher and Visalberghi, 2016; Uher, 2018a, 2021c,d; see Footnote 39). http://researchonindividuals.org.

4 Different terms can denote the same concept (jangle fallacies; Kelley, 1927) and the same term can denote different concepts (jingle fallacies; Thorndike, 1903).

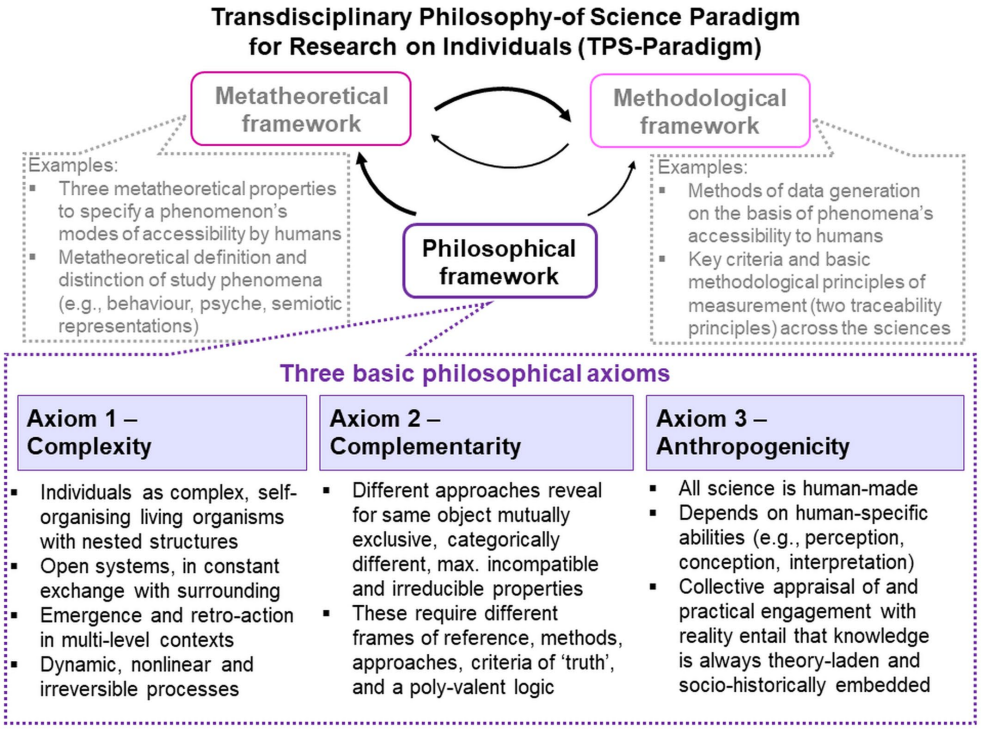


FIGURE 1
The TPS-Paradigm with its interrelated frameworks and three basic philosophical axioms.

been adapted, amongst others, to the body–mind problem (called psyche–physicality problem in the TPS-Paradigm). In this problem, complementarity takes a metaphysically neutral stance making assumptions of neither ontological dualism nor monism whilst emphasising the necessity for *epistemological* and *methodical* dualism to account for the observation of two categorically different realities that require different frames of reference, approaches and methods. This involves a trivalent or even polyvalent (three- or multi-valued) logic rather than a bivalent (two-valued) logic that many psychologists still (implicitly) apply—a hidden remnant of Cartesian thinking (Bohr, 1937; Fahrenberg, 1979, 2013; Walach and Römer, 2011; Walach, 2013; Uher, 2015c).

Axiom 3 – Anthropogenicity: All science is made by humans and thus depends on human-specific abilities

All science is anthropogenic (human-made). Our possibilities to explore and understand the ontological reality in which we have evolved as a species over millions of years are inextricably entwined with and limited by our human-specific perceptual (Wundt, 1907) and conceptual abilities (e.g., interpretations; Peirce, 1958, CP 2.308). Our knowledge about reality is created on the basis of our practical engagement with and collective appraisal of this reality, and is therefore inherently theory-laden, socially embedded and historically contingent (Fleck, 1935; Kuhn, 1962; Valsiner, 2012).

Researchers of individuals face particular challenges because they are individuals themselves and thus not independent from their objects of research. Researchers' own particular positioning in the world—as human beings, members of particular socio-cultural communities, and individuals—makes them insiders in some regards and outsiders in others. This entails risks for *anthropo-centric, ethno-centric and ego-centric biases* that may (unintentionally) influence their scholarly thinking (James, 1890; Fleck, 1935; Weber, 1949; Danziger, 1997), such as when researchers misattribute properties of their own ingroup to outgroups or overlook outgroup properties uncommon in their ingroup. Such *type-I* and *type-II biases* can influence research on both metatheoretical and methodological levels (e.g., choice of research questions, what constitutes data, analytical approaches or interpretational perspectives taken; Uher, 2013, 2015c, 2020a) and are therefore difficult to recognise.

Anthropogenicity highlights a key challenge for psychologists—the distinction of their study phenomena from their means for exploring these phenomena.

Psychologists' cardinal error: Conflating the study phenomena with the means of their exploration—the psych*ical* with the psych*ological*

Key scientific activities such as categorising, generalising, conceptualising, abstracting and analysing are abilities of the human mind. Empirical research is experience-based by definition (from

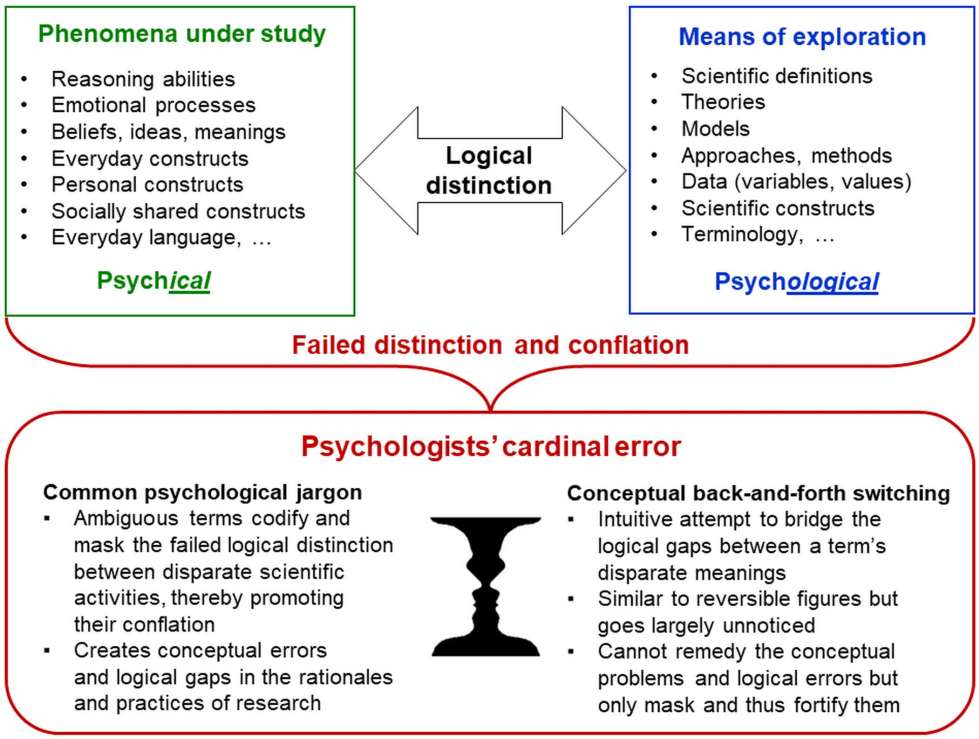


FIGURE 2
Psychologists' cardinal error. Abilities of the human mind are essential for any science; empirical research is experience-based by definition. This complicates for psychologists the logical distinction of their study phenomena from their means of exploring these phenomena, which are therefore often conflated—*psychologists' cardinal error*. Researchers' intuitive conceptual back-and-forth switching between the different meanings that are being conflated masks the logical gaps created and hinders the recognition of these conceptual errors.

Greek *empeiria* for experience). For psychologists—as scientists exploring minds and experience—this complicates the logical distinction of, on the one hand, their study phenomena (e.g., experiences, reasoning abilities, everyday constructs) from their means of exploring these phenomena (e.g., terms, data, methods, scientific constructs) on the other (Axiom 3). In the TPS-Paradigm, this key distinction is emphasised by naming the phenomena of the psyche[5] in themselves as 'psych*ical*' (e.g., mental) and the means of their exploration as 'psych*ological*' (from Greek *-logia* for body of knowledge), as in many non-English languages (Figure 2; Lewin, 1936; Uher, 2016a, 2021b). For example, this article explores psychological problems—professional problems of the scientific discipline—but not psychical problems, which are problems of individuals' mental health. Naming both[6] as 'psychological' cannot reflect this important difference.

Failure to distinguish the study phenomena from the means of their exploration—here called ***psychologists' cardinal error***—is reflected in many practices and jargon established in psychology. It entails conceptual conflations of disparate scientific activities, which create logical gaps that researchers' intuitive conceptual back-and-forth switching between the different activities that are being conflated can only mask but not solve (Figure 2). This logical error has serious implications for entire research programmes because it makes the distinction of disparate elements of research technically impossible, thereby distorting basic conceptions and procedures of science.

Rating 'scales' build on a dense network of 12 conceptual problem complexes

Psychologists' cardinal error is implemented in rating 'scales' in numerous ways—12 metatheoretical and methodological problem complexes are analysed in this section. These problem complexes are tightly interwoven, forming a dense network (Figure 3) that underlies current research practices, which therefore appear to be built on a coherent framework for empirical research. But this coherence masks the faulty assumptions, conceptual problems and logical errors on which these practices are based. This makes these problem complexes so difficult to be detected by those (unwittingly) relying on them. They guide

5 The term psyche is conceived more broadly than mind, thus comprising non-mental phenomena as well.

6 Analogously, we get vir*al* (but not vir*ological*) infections but we do vir*ological* research.

FIGURE 3
Network of 12 conceptual problem complexes (§1–§12) underlying rating 'scales'.

researchers' activities always to the same problematic practices (in different guises), thereby contributing to their perpetuation and psychology's continued crises.

Problem complex §1. Psychologists' own role in their research: Unintended influences

First challenges arise because psychologists are not independent of their objects of research (Axiom 3) whilst, at the same time, essential differences exist between psychologists as researchers and the individuals they research. Specifically, psychical phenomena are accessible (at least partly) only by each individual itself and fundamentally inaccessible by others (Locke, 1999; Uher, 2015a). Therefore, researchers and researchees hold on these phenomena inherently different perspectives, which cannot be shared, enabling observations that the respective other cannot make and may not even be aware of. This disparity can lead researchers of psychical phenomena to mistake their own standpoints for those of the phenomena researched—psychologists' cardinal error. This entails several fallacies to which psychologists are prone (James, 1890).

Six fallacious assumptions are central (Problem complexes *§1a–f*). Researchers often take for granted that the researchees' psychical phenomena are similar to their own, thereby attributing onto them their own beliefs about these phenomena rather than investigating these phenomena as they appear to the researchees. This (*§1a*) *intersubjective confusion* may entail an (*§1b*) *attribution of reflectiveness* when researchers assume that phenomena conscious to themselves are also conscious to the researchees, ignoring that psychical phenomena need not include reflective self-awareness whilst they occur. Moreover, researchers study only fragments of their researchees' psychical phenomena as these are relevant to their research questions, (*§1c*) *ignoring* these phenomena's relevance within the *researchees' horizon of their lifeworld*. It is also fallacious to attribute features of psychological theories to the researchees' psychical phenomena by assuming these are to be understood in terms of categories readily available to researchers, thereby (*§1d*) *substituting knowledge for psychical phenomena*. This also reflects a (*§1e*) *preference of a scientific account over that of the researchees*, which may arise from the researchers' confidence in their privileged position as experts of psychical phenomena generally but overrides the researchees' views who hold the

exclusive access to the particular phenomena studied. The (*§1f*) *misleading availability of ordinary words* makes researchers prone to suppose a substantive entity existing beyond the phenomenon denoted by a word, ignoring that any psychical phenomenon includes much wider ramifications and connotations than words may suggest but also overlooking phenomena not familiarly recognised in language (see Problem complexes §7 Constructification and §8 Naïve use of language-based methods; Ashworth, 2009; Valsiner, 2017a).

Rating methods involve all these fallacies. Ratings are requested on demand, no matter whether or not raters think of themselves as described, consciously reflect on or actually experience the phenomena described. This reflects an erroneous attribution of reflectiveness (§1b) and researchers' focus on the described phenomena's relevance to their own research questions rather than to the researchees' lifeworld (§1c). Item contents are predetermined by the knowledge underlying the theories, concepts and methods that researchers apply for 'scale' development (e.g., for item selection and reduction; McKennell, 1974; Uher, 2015d, 2018a). These practices conflate and (partially) substitute the study phenomena with knowledge that is unrelated to them (§1d). Hence, it is not surprising that item wordings of popular 'personality' 'scales', even if derived from the person-descriptive words of everyday language[7] (John et al., 1988), are actually not amongst those used most frequently in everyday life as is often assumed (Roivainen, 2013; Uher, 2013). Rating items are worded as the researchers understand them given their (pre-) scientific knowledge, whereas raters are not allowed to express their views in their own words, reflecting researchers' preference of a scientific account over that of the researchees (§1e). Item 'scales' are presented without much explanation because researchers take it for granted that raters' understanding of these 'scales' is similar to their own. This ignores substantial, context-dependent variations in raters' item interpretation and use (e.g., Schwarz et al., 1991; Schwarz, 1999; Biesanz and Human, 2010; Rosenbaum and Valsiner, 2011; Lundmann and Villadsen, 2016; Uher and Visalberghi, 2016; Uher, 2018a), reflecting researchers' intersubjective confusion (§1a) and naïve views on language (§1c), which are further explored below (Problem complex §8 Naïve use of language-based methods).

Problem complex §2. Beliefs in researchers' objectivity: Illusions of scholarly distance

These fallacies notwithstanding, and by virtue of their privileged position as investigators knowing (or at least aiming to know) more about the study phenomena than the individuals experiencing them, psychologists typically view themselves as distanced from the individuals and phenomena under study. This disparity, expressed by the term 'participant' ('subject'), creates the illusion of a clear distinction between researcher and researchee, observer and observed. Beliefs in psychologists' objective, uninvolved stance towards their objects of research are rooted in Cartesian thinking (Westerman, 2014) and natural-science research and became established with the introduction of experiments (Danziger, 1985b).

Wundtian scholars still regarded the participant's role as source of information more important than the experimenter's status as operator and attendant and considered both roles as interchangeable (also taking on both roles themselves). This changed fundamentally when Parisian scholars used experiments to study psychopathology (e.g., using hypnosis), which entailed a rigid social differentiation between researchers and the individuals researched. American scholars, in turn, implemented less intense and more impersonal experimenter–participant relations when, commissioned by the American military and government, large-scale investigations shifted psychologists' focus away from single individuals towards populations of individuals (e.g., through group testing). This established a fixed asymmetry between researchers and researchees; participants became anonymous and distant (Danziger, 1985b). Paper–pencil tests, requiring just minimal instruction, became a favoured medium—and paved the way for rating methods.

Today's online surveys distance researchers from researchees even further—direct contact is no longer needed, not even administratively. Yet this does not reduce but increase the impact of fallacious assumptions (Problem complex §1 Psychologists' own role in their research) and of ethno-centric and ego-centric biases (Axiom 3). Problematic findings are therefore not astounding, such as those from international survey panels involving popular 'personality' 'scales'. Instead of showing empirical interrelations as required for psychometric 'scales' (see Problem complex §10 Quantificationism), ratings on items used for the *same* 'personality' construct (e.g., "is generally trusting" and "tends to find fault with others" for 'agreeableness'), varied unsystematically, averaging zero across 25 countries (Ludeke and Larsen, 2017). These and further problematic findings (e.g., incongruent factor structures between counties or between different within-country cohorts) challenge these questionnaires' reliability and validity not only outside of Western, educated, industrialised, rich and democratic (WEIRD) populations but also their adequacy and predictive utility for studying individuals within Western countries (Hanel and Vione, 2016; Laajaj et al., 2019; Condon et al., 2021).

Maximising scholarly distance alienates psychologists from the psychical phenomena under study, which are inherently accessible only to the researchees. Lack of direct contact impedes testing researchers' (e.g., ethno-centric and ego-centric) assumptions and interpretations, and thus implementation of any corrective means as well as discussions about what objectivity,

7 In lexical approaches, the person-descriptive words in a language's lexicon are used to categorise the individual differences considered to be most important in a sociolinguistic community (Allport and Odbert, 1936; Uher, 2015d).

given the peculiarities of psychical phenomena, could actually mean.

Problem complex §3. Mistaken dualistic views: Individuals as closed systems

Psychologists' beliefs in their own objectivity, by virtue of being researchers, entail further problematic conceptions. Specifically, seen from the researchers' own—supposedly objective—observer standpoint, individuals are often conceived as opposed to, thus separate from the conditions in which they are being studied, as reflected in behaviourist input–output models but also in statistical independent variable–dependent variable (IV–DV) models, amongst others. Individuals are seen as reacting to standardised stimuli that are thought to have the same meaning for all individuals. This conceptual separation underlies, for example, person–situation and person–environment (nature–nurture) research (e.g., in trait psychology) whereby the researchers determine what constitutes a 'stimulus', 'situation' or 'environment', etc. and what meanings these may have for the researchees.

Such dualistic, researcher-determined views reflect a simplistic thinking that facilitates researchers' work and that enables flexible adaptations to the given knowledge applied (Valsiner, 2017b). But it falls prone to the biases and fallacies on the researchers' part (Axiom 3; Problem complex §1 Psychologists' own role in their research). It also overlooks that, as complex open systems (Axiom 1), individuals are interrelated with only those parts of their external surrounding for which they are receptive and with which they can interact given their species-specific, community-specific and individual-specific abilities. Thus, what constitutes an individual's external context (e.g., a 'situation') are just parts of its entire external surrounding.[8] This external context (from Latin *con* + *texere* for woven together) is defined by characteristics of that individual with which these parts are functionally interrelated (e.g., through its perception or conception of them). Hence, an individual's external context cannot be conceived independently of that individual although it is—like all parts of the individual's external surrounding—physically located outside of the individual. The specifics, functions and meaning that an external context has for a given individual may therefore not be apparent for others (e.g., researchers; von Uexküll, 1909; Valsiner, 1997; Uher, 2015a; see also Rotter's, 1954, 'psychological situation' concept).

Psychologists' common consideration of rating items as (verbal) stimuli to which researchees respond are prime examples of such dualistic concepts. Depending on the researchers' theoretical views, raters' responses are used to explore, for example, either individuals' characteristics (in trait psychology), cultural influences (in cross-cultural psychology), or relative gene versus environment influences (in quantitative genetics). Thus, raters' responses are flexibly attributed different meanings as needed to match researchers' theories and to answer their particular questions (see analogously, Bandura, 1996)—reflecting the psychologist's fallacies (Problem complex §1 Psychologists' own role in their research). To study individuals' systemic interdependences with their contexts, it requires *inclusive concepts*, in which relevant parts of an individual's surrounding—despite their physical independence from the researchee as seen from the researcher's observer perspective—are identified and conceived only in *dependence* of the researchee's individual characteristics (see Problem complex §12 Nomotheticism; Uher, 2015a,c; Valsiner, 1997, 2017b).

Problem complex §4. Lack of definition and theoretical distinction of study phenomena: Conceptual conflations and intersubjective confusions

Psychology's core constructs (e.g., mind, behaviour, actions) are poorly defined; common definitions are discordant, ambiguous, overlapping and circular (Zagaria et al., 2020). At the same time, terms and constructs for specific psychical and behavioural phenomena proliferate chaotically, creating countless jingle–jangle fallacies (Uher, 2021b). These problems reflect many of the psychologist's fallacies (Problem complex §1 Psychologists' own role in their research)—and analogous ones. Specifically, as all socialised persons, psychologists have a complex everyday knowledge, which includes pre-scientific concepts and words for many of their study phenomena (Axiom 3; Uher, 2013). These may be helpful to get research started until more elaborated concepts, terms and definitions are developed. But when researchers substitute their pre-scientific knowledge for their study phenomena (Problem complex §1d) and attribute their own understandings onto their colleagues ("we all know what they mean"; Problem complex §1a), conceptual advancements are hampered.

In the TPS-Paradigm, behavioural and psychical phenomena[9] are metatheoretically distinguished from one another without

8 von Uexküll (1909) differentiated Umwelt (external context) as opposed to Aussenwelt (external surrounding).

9 In the TPS-Paradigm, a *phenomenon* is defined as anything humans can perceive or (technically) make perceivable and/ or that humans can conceive of (Axiom 3; Uher, 2015c). This definition differs from previous definitions, such as Kant's (1781/1998) phenomena–noumena distinction. Unlike the latter, it considers that appearances are never purely perceived as such but influenced by the perceivers' (pre)conceptions. Any attempt for making such distinctions for epistemological purposes is affected by the same problem. What humans cannot conceive of, cannot be researched (Axiom 3; see Wittgenstein, 1922, on linguistic prerequisites). This is particularly important for research on individuals where many phenomena are studied directly without using technical instruments (Uher, 2019).

implying their ontological separability (Axiom 2). *Behaviours* are defined as the "external changes or activities of living organisms that are functionally mediated by other external phenomena in the present moment" (Uher, 2016b, p. 490). The *psyche* is defined as "the entirety of the phenomena of the immediate [non-mediated[10]] experiential reality both conscious and non-conscious of living organisms" (Uher, 2016a, p. 303). These definitions highlight important points for research on individuals and the present analyses of rating 'scales'. Behaviours are publicly accessible and physically describable, psychical phenomena are not. Psychical phenomena can be inferred from behaviours but neither are psychical phenomena systematically related to behaviours nor are they contained in the behaviours themselves. Most behaviours are ambiguous because they simultaneously possess various features and can therefore be interpreted differently regarding any possibly associated or causally underlying psychical phenomena (e.g., intentions, goals, feelings; Shweder, 1977; Smedslund, 2004; Toomela, 2008; Uher, 2015c).

A key point behind these metatheoretical definitions is the distinction between *description* versus *explanation*. Behaviours can be described in their momentary physical properties and situational locatedness. But their explanations can go well beyond the here-and-now and can invoke many different *interpretive perspectives*, which all follow logical principles (Kelly, 1955; Smedslund, 2004) yet without being logically determined by a behaviour itself (Shweder, 1977). The same physically described behaviour can be interpreted very differently, depending on the contexts that the interpreting individuals consider for themselves as observers and the individual observed (if applicable). Hence, a behaviour can have different *meanings*, each involving different explanations—which may all be logically justified and thus appear to be reasonable—but of which only some apply in a given case.

The metatheoretical distinction between psychical and behavioural phenomena is also important to explore their relations with one another and with other phenomena internal and external to individuals' bodies—such as in actions. *Actions* are conceived in the TPS-Paradigm as complex kinds of phenomena comprising (a) behaviours (external changes and activities), their relations with (b) the phenomena of the individual's external context that are mediating the behaviours' functionality in the present (see behaviour definition above), and with (c) psychical phenomena directing and controlling these behaviours and their outcomes (e.g., intentionality, goal-directedness). From a certain level of psychical complexity (Axiom 1), individuals are able to conceive of and evaluate the outcomes of their own behaviours and to use these concepts to anticipate possible future outcomes. This enables individuals to adapt their own behaviours, to plan ahead and develop intentions. Anticipated future outcomes can motivate, guide and regulate individuals' current and future behaviours (Kelly, 1955; Valsiner, 1998; Searle, 2003; Smedslund, 2004; Bandura, 2006; Uher, 2013, 2015a,c, 2018b; Tomasello, 2014). Hence, actions are far more complex and involve more diverse kinds of phenomena than just behaviours.

In common psychological jargon, behavioural and psychical phenomena are often conflated, such as when naming both as ('inner' and 'outer') 'behaviours'. This blurs description with interpretation and explanation, opening doors to inferential fallacies and attributional biases (Table 1). It also entails that behavioural and psychical phenomena are methodically treated the same, ignoring their different modes of accessibility, which require different research methods (Axioms 2 and 3; Uher, 2019). This frequent conflation may be an attempt to overcome behaviourist ideas and Cartesian ontological dualism.[11] It may also derive from everyday experience as everyone can notice the tight links between their own psychical and behavioural phenomena. Everyday language is full of concepts and terms intermingling descriptions with explanations of behaviour. Normal adults often talk as if they had observed others' psychical phenomena (e.g., intentions)—although causal inferences can be made only on the basis of premises; but these premises often remain implicit and can logically justify alternative inferences as well (Shweder, 1977; Smedslund, 2004).

The use of everyday language incorporates these fallacies into rating 'scales'. Items rarely describe observable behaviours only, such as using descriptive action verbs (e.g., talk). Most items are inferential, such as through trait-adjectives (e.g., jealous), trait-nouns (e.g., opportunist), state verbs (e.g., envy) or interpretive action verbs (e.g., help; Semin and Greenslade, 1985; Semin and Fiedler, 1988). That is, items may require raters to judge phenomena that are actually imperceptible to them (e.g., others' emotions) or no longer perceivable (e.g., past occurrences to judge habitual behaviours as in 'personality' ratings). Inferential and complex wordings do not preclude research, as interpretive analyses of textual materials demonstrate (Fahrenberg, 2002, 2003). But in ratings, they obscure which specific phenomena and which specific aspects of them raters actually consider (Problem complexes §1 Psychologists' own role in their research, §8 Naïve use of language-based methods and §10 Quantificationism).

Problem complex §5. Reductionism: Category mistakes, atomistic fallacy and decontextualisation

The interpretation of rating-based findings as reflecting "psycho-physical mechanisms" underlying individuals' behaviour (common, e.g., in trait psychology) reflects further

10 Immediacy indicates absence of phenomena mediating their perception (Wundt, 1896).

11 The TPS-Paradigm's concepts have also been misunderstood at times as either behaviourist or ontologically dualist (e.g., by Franz, 2021), ignoring their philosophical and theoretical foundation in theories of complex dynamic systems and of complementary relations (Axioms 1, 2). The present improved description may hopefully clarify these misunderstandings.

TABLE 1 Common conflations in psychology.

'Behaviour', 'Inner behaviour' – 'Outer behaviour'	
Psyche	Accessible only privately; thus phenomena of psychical life, which cannot be directly accessed by others but only by the individual itself; defined as "the entirety of the phenomena of the immediate [non-mediated] experiential reality both conscious and non-conscious of living organisms" (Uher, 2016a, p. 303)
Behaviour	Accessible publicly; thus phenomena that occur external to individuals' bodies; defined as the "external changes or activities of living organisms that are functionally mediated by other external phenomena in the present moment" (Uher, 2016b, p. 490)
Cause – effect	
Cause	Entity providing the generative force that is the origin of something (its effect) and that is thus responsible for bringing about the latter
Effect	Something that is produced by a cause or agent and that is thus a result of these latter; something that follows naturally or logically
Construct – referent	
Construct	Conceptual system that refers to a set of independent entities (construct referents) on more abstract levels and that is constructed by filtering relevant information about these referents (e.g., by (de-)emphasising aspects that are considered relevant); thus, it constitutes a conceptual entity, which is not the same as the referents to which it refers and which does not exist in itself as a concrete entity
Referent/Construct referent	Independent entities that are regarded as meaningfully related in some ways or for some purpose although they actually never occur all at once and that are therefore considered only on more abstract conceptual levels as a joint entity (the construct) and that are thus not the same as and different from that conceptual entity
'Data'	
as Study phenomena	The study phenomena in themselves (located, e.g., in individuals) that are to be observed. Not to be confused with psychometricians' so-called 'observed' or 'manifest' data, which refer to the raw data and are thus sign systems encoding information about the study phenomena but they are not these phenomena in themselves
as Sign systems	Variables and values (located, e.g., on spreadsheets) carrying information about the study phenomena that are to be analysed. Psychometricians commonly refer to the raw data as 'observed' or 'manifest' data and the modelled data results as 'latent' data; all these data are sign systems but neither the study phenomena in themselves nor structures underlying these phenomena
Description – explanation	
Description	Factual statements and discourse intended to give an account of the characteristics of something experienced
Explanation	Statements and discourse that conclusively derive something unknown from known elements such that its origin and development can be recapitulated and made comprehensible
Individual differences – Individuality/'personality'	
Individual differences	Differential patterns describing differences between individuals in a sample (e.g., analysed using variable-oriented approaches); they characterise the population and cannot inform about the single individuals
Individuality/'personality'	Individual-specific patterns (Uher, 2013) describing an individual's peculiarities, which implies differences from others over some time but analysed on the individual level (e.g., using person—/individual-oriented approaches) and thus characterising the single individuals
Methodology – method	
Methodology	Philosophy and theory of the approaches (ways) and methods suited to explore particular study phenomena
Method	Specific practices, procedures and techniques that are used to perform the operations that are necessary for the investigation, manipulation or elicitation of study phenomena
Numeral – number (special case of signifier–referent conflation)	
Numeral	Signifier, sign vehicle, written or spoken entity (e.g., graphemes or phonemes) often used to indicate numbers (quantitative properties) but also letters or just categorical (non-quantitative) information (e.g., phone or house 'numbers')
Number	Arithmetical value, mathematical entity describing a quantity
'Operational definition'	
Definition	Formal description of the nature, properties or essential qualities of something
Operationalisation (proceduralism)	Reporting design and method details used for empirical investigation

(Continued)

TABLE 1 (Continued)

Psychical – Psychological	
Psychical	Phenomena of the psyche in themselves (e.g., mental, emotional, cognitive, experiential etc.)
Psychological	Means used to explore psychical phenomena and the body of knowledge developed about these phenomena (from Greek -logia for body of knowledge)
Signifier – meaning	
Signifier	Sign vehicle, written or spoken entity (e.g., graphemes or phonemes) used to represent a referent and its meaning; therefore publicly accessible
Meaning	Sense, purpose, significance, intent or definition that something (e.g., a word, action, or concept) has for somebody
Signifier – referent	
Signifier	Sign vehicle, written or spoken entity (e.g., graphemes or phonemes) used to represent a referent and its meaning; therefore publicly accessible
Referent/Sign referent	Designatum, what is being designated and referred to by the signifier of a sign whether concretely perceivable, conceived, imagined or fantasised (e.g., objects, events, concepts)
'Variable', Variable – referent	
Variables/Data variables	Sign systems encoding information about the study phenomena for the purpose of recording and analysing this information on symbolic levels in lieu of the actual study phenomena (e.g., using statistics). Confusingly, psychometricians commonly refer to the raw data variables as 'observed' or 'manifest' variables and the modelled results as 'latent' variables; but all these variables are sign systems and neither the study phenomena in themselves nor structures underlying them
Variable referents	Study phenomena about which information is being recorded and explored by means of sign systems

Definitions used in the TPS-Paradigm for Research on Individuals to clarify terms and concepts and to avoid jingle–jangle fallacies and conceptual conflations.

fallacies—those of *reductionism*. Reduction itself is basic to any science. Approaches and methods to reduce study phenomena, their relations, data, etc. are fundamental for scientific model development (e.g., reduction approaches in taxonomic individual differences research; Uher, 2015d, 2018b). By definition, models are reduced (less detailed) representations of complex parts of reality (Axioms 1 and 3). Hence, not all reductions are wrong; but some are fallacious (Fahrenberg, 2013)—here called *reductionism*.

Three forms of reductionism—ontological, epistemological and methodological—are common. *Ontological reductionism* refers to claims about the relations between phenomena whereby it is assumed that complex phenomena can be described in terms of simpler, more fundamental ones. An example is the idea that psychical phenomena would constitute just neuronal firing through electric impulses and neurotransmitters (Brigandt and Love, 2022).

Epistemological reductionism is the claim that knowledge about one scientific domain, typically about higher-level phenomena, can be reduced to another body of scientific knowledge, typically about a lower or more fundamental level. An example is to assume that higher-level phenomena could be explained by lower-level phenomena, such as psychical phenomena by underlying neurophysiological phenomena. But Wundt (1902-1903) already highlighted that, even if brain processes would be as clear to us as clockwork, this could not elucidate the interrelations of psychical phenomena in themselves. Analogously, when asking about an object's weight, the statement "it is red" provides no answer—this constitutes a *category mistake*. Mass and colour are different categories and belong to different systems of description; weight cannot be expressed in terms of colour. The popularity of analogous statements about psychical and physical (e.g., neurophysiological) phenomena does not make them any more true. These phenomena are complementary—both are needed for comprehensive accounts of individuals but one cannot be reduced to or transformed into the other. Such attempts constitute a category mistake. Psychical and neurophysiological processes require different frames of reference, systems of description, epistemological principles and perspectives, which cannot be reduced to each other (Axiom 2; Walach and Römer, 2011; Fahrenberg, 2013; Walach, 2013). For empirical research, scientists may focus on just some kinds of phenomena, thereby blanking out others, such as neurophysiologists and cultural psychologists do. But holistic accounts of individuals always require knowledge of all the different kinds of phenomena occurring in (relation to) them.

Methodological reductionism is the claim that complex systems are most fruitfully investigated at the lowest possible level and could thus be understood by dissecting them into their supposedly isolable building blocks. Such *mechanistic* and *elementarist* views may be useful to explore invariant physical phenomena. But studying elements regardless of their interrelations with other elements and of the contexts in which they occur, meets its limits in living systems (Axiom 1). Knowledge about a cell's *decontextualised* parts and biochemical

components does not reveal how they function together in the intact living cell (Rothschuh, 1963; Brigandt and Love, 2022). Elementarist reductionism reflects the *atomistic fallacy* whereby, from information obtained at lower levels, incorrect inferences are made at higher levels of organisation (Diez Roux, 2002). In psychology, elementarist reductionism is reflected, for example, in Western psychologists' categorisations of psychical phenomena into memory, perception, motivation, emotion, language etc. and their treatment as separate processes—assuming their study in isolation from their contexts could still reveal meaningful information about their functioning in the individual (Danziger, 1990). Elementarism allows researchers to explore only problems of structure, which are analytic and descriptive, but not problems of process and functioning, which are causal (Bartlett, 1932/1995).

Rating items build on the atomistic fallacy. They are seen as manageable chunks of information that could be understood in isolation from the contexts in which they are used—such as the other items, the raters interpreting them, the specific phenomena that raters decide to judge, those they may consider for comparison, the explanatory perspectives they take on them, etc. Decontextualisation could hardly be any more radical. These decontextualised chunks of information are then put together again using statistical procedures, thus using knowledge that is unrelated to the study phenomena described in the items and the contexts of their use (Problem complexes §1d Psychologists' own role in their research; §3 Mistaken dualistic views and §11 Statisticism). The popular interpretation of statistically reduced rating data as reflecting "psycho-physical mechanisms" that are heritable, universal and evolutionarily adaptive (e.g., 'traits'; McCrae and Costa, 2008; Buss, 2009) builds on several reductionist fallacies.

Elementarist reductionism is tightly linked to operationalism.

Problem complex §6. Operationalism: Logical errors and impeded theory development

Wundt developed substantial theoretical and conceptual frameworks for an enormous breadth of psychical phenomena, ranging from psychophysics to cultural psychology (Völkerpsychologie). But many of his concepts were too sophisticated for extensive empirical investigations (Danziger, 1979; Fahrenberg, 2019). Behaviourists, in turn, rigidly avoided altogether to conceptualise phenomena that are inaccessible in others. To establish the fledging discipline's empirical research on its primary study phenomena, psychologists turned to operationalism from physics (Bridgman, 1927) and adapted it to their purposes in their own specific ways (Feest, 2005). Operationalism seemed to offer a solution, enabling both empirical research and concept development in a surprisingly straightforward manner.

> Operationism consists simply in referring any concept for its definition to the concrete operations by which knowledge of the thing in question is had (Stevens, 1935, p. 323).

Its strong links to logical positivism and statistical advancements like factor analysis (Spearman, 1904; Thurstone, 1937) have firmly anchored empiricism in psychology's methodological conventions. Still today, operationalism is considered an essential feature of rigorous psychological research (AERA, APA, and NCME, 2014).

But operationalism, both as introduced in physics and its psychological variants, has been fundamentally criticised in its most basic assumptions (e.g., Benjamin, 1955; Wallach, 1971; Bickhard, 2001; Feest, 2005). The idea that a study phenomenon's meaning could be established through the operations needed for its investigation, manipulation or elicitation conflates the study phenomena with the means of their investigation—psychologists' cardinal error. Specifying operational procedures may help piloting conceptual work about a study phenomenon. But ultimately, operational specifications must be replaced by proper theoretical definitions (Green, 2001; Feest, 2005)—otherwise, this leads to further logical errors. For example, when reasoning ability is operationally 'defined' as test performance, this ability cannot also be used to explain this performance. A phenomenon cannot be defined by its effects; such assumptions *conflate cause with effect* (Table 1; Hibberd, 2019).

Moreover, if a construct's definition depends on a specific procedure, even if just partially, then every change in procedure defines a new concept. This reasoning may have contributed to the *proliferation of constructs* because psychologists tend to disagree much less on their findings than on their construct operationalisations and therefore prefer to use each their own, leading to overlapping constructs and countless jingle–jangle fallacies (Feest, 2005; Uher, 2021b). Yet the idea that every procedural change also defines a new concept contradicts all sciences' striving to advance their portfolio of methods, including those suitable for studying well-known phenomena (Wallach, 1971). It also contradicts the realist ontology that many proponents of operationalism assume for psychical phenomena (Hibberd, 2019). Psychologists hoped to solve this problem with "convergent operationalism," which involves multiple independent operations for the same construct (Campbell and Fiske, 1959). But linking constructs with classes of operational procedures does not solve the basic problem that two disparate scientific practices—(1) reporting design and method details, and (2) defining the concept or study phenomenon—are being conflated (Table 1; Green, 2001; Hibberd, 2019).

Rating methods are prime examples of operationalism. As verbal materials, items can be easily reworded and redesigned so that new rating instruments can be created at libitum and low cost—and with them new constructs (Axiom 3). The verbal provision of rating 'scales' and brief instructions to raters greatly facilitates the documentation of the operational procedures that are used to specify given constructs (Uher, 2015d). Together with

advanced statistical methods and facilitated by their computerised implementation, rating 'scales' have therefore become for many psychologists the preferred tool allowing them to conduct empirical research on almost any topic (Lamiell, 2013; Maul, 2017; Arnulf, 2020).

This a-theoretical instrumentalism entails the belief that any rating 'scale' that is nominally associated with a study phenomenon could be a valid method for its investigation (e.g., 'extraversion scale', 'neuroticism scale'). This *toolbox thinking* contributes to the proliferation of substantially overlapping constructs and their pertinent, likewise overlapping rating 'scales' (e.g., there are dozens of depression and anxiety 'scales'; Sechrest et al., 1996). Toolbox thinking invites researchers to choose their topics and questions by the methods available rather than vice versa, thereby enacting worlds that are fit for their methods (Law, 2004). But semantics do not define methods. This practice indicates a failure to reflect on and to use language-based methods (Problem complex §8 Naïve use of language-based methods), and contributes to the persistence of problematic research practices and crises in psychology (Toomela, 2009; Valsiner, 2017b).

The common term 'operational definition' misleads researchers to assume that mere descriptions of operational procedures could substitute for the theoretical work on a study phenomenon or concept (Table 1), such as when the results of rating operations are used to define or even "identify" constructs—as often done with factor-analysed rating data (e.g., in differential psychology). But scientific definition is logically prior to the scientific task of empirical investigation (Hibberd, 2019). *Operationalisation* in itself, however, is unobjectionable and even needed for construct research (Problem complex §7 Constructification)—as long as it is meant only as specifying the operations or procedures used to investigate (e.g., elicit, test, assess) a research object. In nomological networks, for example, psychologists define the target construct and its sub-constructs in a theoretical framework, specify the operations needed for its investigation in a separate empirical framework and systematically link both frameworks (Wittmann, 1988).

But this *procedurism* is not scientific definition. It is illogic to treat procedurism as constitutive for a phenomenon's definition or a concept's meaning (Hibberd, 2019). It also conflates theories about the study phenomena with theories about methods (Uher, 2021c). Theories about study phenomena are tested *via* predictions that can be derived from these theories; this does not require operationalisation. A concept's theoretical meaning, the testing of hypotheses and theories, and the procedures of measurement or other empirical investigation are not identical. Operationalism conflates these disparate scientific activities, making their distinctions technically impossible and distorting conceptions and procedures of science. "This contributes to the lack of understanding of theory in psychology and to the relative naivety of the theoretical work that exists in psychology" (Bickhard, 2001, p. 42).

Psychologists commonly discuss operationalisation with regard to constructs.

Problem complex §7. Constructification: Studying constructs without also studying their intended referents

Constructs are central to psychology (Maraun et al., 2009). But research on constructs is plagued by their vague, inconsistent and contradictory definition and use (Lovasz and Slaney, 2013; Slaney, 2017), leaving many psychologists utterly confused:

> We do not know what constructs are, that is, we have rarely come across a clear description of what something should be like in order to deserve the label 'construct'. Constructs, as far as we are concerned, are truly shrouded in mystery, and not in the good old scientific sense that we currently don't know what they are, but will know when we're finished doing the relevant research, but in the sense that we don't really know what we are talking about in the first place (Borsboom et al., 2009, p. 150).

The main source of this confusion is that constructs are sometimes interpreted as theoretical concepts and sometimes as the study phenomena denoted by such concepts and that both interpretations are often conflated (Danziger, 1997; Slaney and Garcia, 2015; Uher, 2021c,d)—psychologists' cardinal error.

Constructs, like all concepts,[12] are products of the human mind as are ideas, theories and knowledge (Axiom 3). Thus, concepts are psychical phenomena; this is their ontology—a fact that other disciplines can conveniently ignore and therefore oppose 'natural' and 'real' phenomena to the concepts designating them. But are psychical phenomena—is the human mind unnatural and not real (surreal)? What difference could there be, ontologically, between scientific constructs and the constructs that people develop in everyday life (see Kelly, 1955)? They may differ in complexity, coherence of linguistic codification and use. But both can only be thought and conceived by persons—and both even by the same person. The difference between them is thus not in kind; a distinction is made only for epistemological purposes. Constructs do not exist outside of the realm of psychical phenomena (Axiom 3)—a first challenge inviting psychologists' cardinal error (Uher, 2021c,d).

A construct is a *conceptual system* that *refers to* a set of entities—the construct *referents*—that are regarded as meaningfully related in some ways or for some purpose *although they actually never occur all at once* and that are therefore considered only on more abstract levels as a joint entity. That is, constructs do not exist as concrete entities in themselves; they are only thought of as entities—they are *conceptual entities*. For example, the construct 'intelligence' may refer to the entirety of a person's problem-solving abilities, but these abilities can never be observed all at once. The referents of the

12 Constructs and concepts are both abstract ideas. Constructs tend to have more heterogeneous referents and therefore to be more abstract and complex. But attempts to clearly differentiate them are ultimately arbitrary.

construct 'climate'—an area's long-term weather conditions—cannot be observed all at once either.

The conceptual nature of constructs has three important implications. First, researchers can empirically study only a (manageable) subset of a construct's referents that they choose to serve as construct *indicators* (e.g., test items), whereas the (hypothetical) universe of a construct's referents and their essential features form the basis of its theoretical definition. This highlights again the difference between construct definition and operationalisation (Problem complex §6 Operationalism). Second, as conceptual entities, *constructs can refer to entities of all kinds* (e.g., abiotic, biotic, psychical, social, cultural). Specifically, a construct's referents can involve various entities of the same kind of phenomenon (e.g., various problem-solving abilities in the construct 'intelligence') but also entities of heterogeneous kinds of phenomena (e.g., behavioural, psychical, and physiological phenomena in the construct 'extraversion'). This entails numerous challenges for both conceptual and empirical research.

Conceptually, different kinds of phenomena can be integrated seamlessly into the same construct although they differ in accessibility, thus requiring different methods of investigation (Uher, 2019). Such conceptual conglomerates of heterogeneous referents are *blended constructs* (Uher, 2018b). For example, behaviours are accessible publicly by observation, psychical phenomena only privately through self-observation and self-report, and physiological phenomena require physical measuring devices. A construct's referents may also have forms of occurrence as diverse as discrete objects (e.g., brain structures), instantaneous events (e.g., heart beats) and continuous processes (e.g., thinking). But this does not hinder their conceptual integration into the same construct. Indeed, constructs are indispensable to study processes because, at any moment, only a part of a process exists. Processual phenomena, such as many behavioural and psychical ones, can therefore be conceived only by generalising and abstracting from their occurrences over time (Whitehead, 1929). For this reason, constructs are essential for psychological research (Uher, 2021c,d).

Qualitatively and quantitatively different referents can be conceptually integrated through abstraction. *Conceptual abstraction* allows humans to filter information about complex entities and to reduce their complexity by emphasising some of their aspects and deemphasising others (Whitehead, 1929), depending on their ascribed (ir)relevance for a particular meaning or purpose (e.g., social valence, prediction). For example, to facilitate the distinction between similar individuals or similar experiences, people often exaggerate in their constructs minor differences (e.g., between individuals or groups) that are considered to be socially relevant and that then appear, in people's minds, to be much larger than can actually be observed, thereby acquiring *salience* (Lahlou, 1998; Uher, 2015c).

All humans develop constructs (Axiom 3)—individual theories to describe regularities occurring in their daily lives and to discriminate between experiences they have made, and which are therefore idiosyncratic and personal (Kelly, 1955). Individuals use these *personal constructs* to make predictions, gain cognitive control over future events and guide their own actions. They test their constructs' appropriateness (viability) for these purposes with new experiences, thereby developing their constructs further, integrating and organising them by their level of generality into complex *construct systems* (Axiom 1). Members of the same community, using their socio-culturally shared experiences, can develop an understanding of others' personal constructs and the actions derived from them, enabling joint understanding and coordinated action. Constructs that proved to be viable to predict and control individuals' actions in everyday life—thus, to distinguish between individuals in socially relevant ways and to establish normativity—become *socially shared constructs* (Kelly, 1955) and encoded in natural everyday language (e.g., person-descriptive words; Klages, 1926; Tellegen, 1993).

Constructs can be construed on all levels of abstraction—from referents that are concretely perceivable at a given moment (e.g., specific behavioural acts) over referents that are conceptual and generalised in themselves (e.g., 'sociability') up to referents that are only imagined (e.g., future society) or fantasised (e.g., supernatural beings). That is, constructs can refer also to other constructs representing their content on higher levels of abstraction (e.g., a construct 'sociability' may refer to more specific constructs such as 'gregariousness', 'talkativeness' and 'approachability'). This entails *nested conceptual structures* (symbolised by words) in which meanings and referents can be 'inherited' from the various more specific constructs that they comprise (Figure 4; Uher, 2013, 2021c,d). (For the special role of language therein and their exploration in semantic networks, see Problem complex §8 Naïve use of language-based methods). Constructs and their linguistic labels thus contain complex implicit meanings and conceptual structures (Vygotsky, 1962; Lahlou, 1996). This highlights the third implication of the constructs' abstract conceptual nature. Constructs

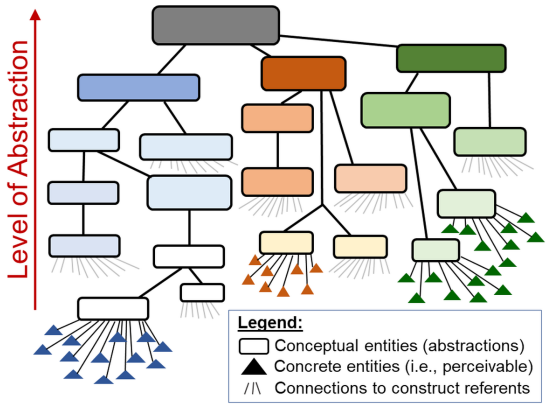


FIGURE 4
Constructs: Conceptual systems with nested structures. Constructs can be construed on all levels of abstraction. They can refer to concrete entities but also to other constructs, representing their contents and thus the referents to which these constructs refer on higher levels of abstraction. This entails nested conceptual structures (symbolised by words) in which meanings and referents can be 'inherited' from the various more specific constructs that they comprise.

imply more (surplus) meaning than the concrete indicators by which they can be empirically studied (Problem complex §6 Operationalism) and that therefore cannot be reflected by a construct in the same ways as individuals can perceive them at any given moment (Vygotsky, 1962).

> There is no present for all the elements and structures of conceptual systems at once (Althusser and Balibar, 1970, p. 318).

Psychologists' frequent confusions around constructs can thus be traced back to two interrelated problems, (1) *lack of conceptual understanding of what constructs actually are* and (2) *failure to distinguish scientific constructs from their referents (and indicators)*. Insufficient conceptual understanding involves a lack of awareness that constructs, as psychical phenomena, can be explored in themselves (e.g., people's everyday constructs in lexical 'personality' research; Tellegen, 1993) but that constructs are also important means of exploration (e.g., scientific constructs about everyday constructs, such as the Big Five 'personality' constructs). That is, disparate elements of research can be constructs—but, in a given study, the same construct logically cannot be both. Ignorance of this important point led to the implementation of the logical errors of operationalism in psychological research (Problem complex §6 Operationalism). A study's scientific constructs (e.g., the Big Five 'personality' constructs) serve as means of exploration (e.g., as categorical summary statements; Wiggins, 1979) and should thus not be mistaken for the study's actual study phenomena (e.g., "universals" of human nature that are "invariant across human cultures," McCrae and Costa, 1997, and thus potentially phylogenetic in origin, McCrae, 2009). Still, scientific constructs can also be explored in themselves—but in other studies using other, higher-order scientific constructs.

These peculiarities make it difficult to distinguish scientific constructs from their referents (and indicators). But constructs only *refer* to particular entities—they *are not* these referents in themselves. Constructs and their linguistic labels (Problem complex §8 Naïve use of language-based methods) facilitate thinking and communication (Vygotsky, 1962). In everyday life, people frequently conflate constructs with their referents (e.g., disease labels taken for illness-causing entities). But *construct–referent conflation* (Table 1) entails serious problems for research because it conflates study means with study phenomena—psychologists' cardinal error. Recognising such conflations is difficult especially when a construct's referents are not directly accessible for researchers or conceptual in themselves, as is mostly the case in psychology. Construct–referent conflation was shown to underly confusions about (a) the interrelations between everyday constructs and scientific constructs (Uher, 2013), (b) construct operationalisation, nomological networks and representation theorems (Uher, 2021d), (c) concepts of latent traits, variables and models in psychometrics (Maraun and Halpin, 2008; Uher, 2021c), and (d) disparate notions of the terms 'hypothetical' (Lovasz and Slaney, 2013) and 'unobservable' (Uher, 2021d) in discussions about constructs.

Rating 'scales' promote construct–referent conflation because items are commonly inferential and refer to (often heterogeneous kinds of) study phenomena in more general and abstract terms (Problem complex §4 Lack of definition and theoretical distinction of study phenomena). Such items describe (blended) constructs. Their contents can be judged on the mere conceptual-semantic level (Problem complex §8 Naïve use of language-based methods; Shweder, 1977; Arnulf et al., 2014). But even if raters consider concrete phenomena (e.g., specific behavioural acts), to produce an overall judgement (e.g., on their intensity), raters must implicitly compare them at least over different occurrences, thus over time if not also over different individuals (and phenomena, e.g., other behaviours). That is, ratings inherently involve retrospective considerations and thus abstraction and generalisation. This means, in turn, that ratings cannot capture specific occurrences of phenomena, which, however, is needed for measurement (Problem complex §10 Quantificationism). To study processual phenomena rather than just constructs *about* them, researchers must record the given phenomena as, when and where they occur and over some time using *nunc-ipsum methods* (from Latin *nunc ipsum* for at this very instant; Uher, 2019), such as methods of Ambulatory [ecological momentary] Assessment (Fahrenberg et al., 2007; Mehl, 2017), and must analyse the data thus-obtained for regularities, structures and relations (Van Geert and Van Dijk, 2002; Molenaar, 2008). But in ratings, the phenomena of interest (e.g., social behaviours, emotions) are typically not even present during data generation (e.g., self-rating on screen). The fact that ratings can be generated regardless clearly shows that they are based on raters' beliefs, ideas and knowledge—and thus reflect personal and socially shared constructs that raters have developed *about* the phenomena described rather than these phenomena in themselves.

Exploring everyday constructs is worthwhile in itself and informative about socio-cultural phenomena. But rating data are often interpreted as reflecting information about individuals' experiences and behaviours *in themselves*, ignoring that these processual phenomena require methods of data generation other than ratings and therefore remain unexplored (Molenaar, 2008; Rosenbaum and Valsiner, 2011; van Geert, 2011; Uher, 2013, 2016a,b). In consequence, researchers develop scientific constructs without also studying their intended referents—here called *constructification*.[13] The popularity of rating 'scales', thought to enable efficient empirical research on a broad range of behavioural and psychical phenomena, has institutionalised this fallacy widely in psychology. Researchers who rely exclusively on rating 'scales' are studying everyday knowledge about their study phenomena, thus laypeople's generalisations and abstractions with all the biases, inconsistencies and inferential fallacies that these are known to contain (Problem complex §4 Lack of definition and

13 The term constructification is used because the term constructionism is already established for constructivist approaches in psychology and social sciences. Constructionism refers to an epistemology (e.g., Gergen, 1985); whereas constructification refers to the fallacy described above.

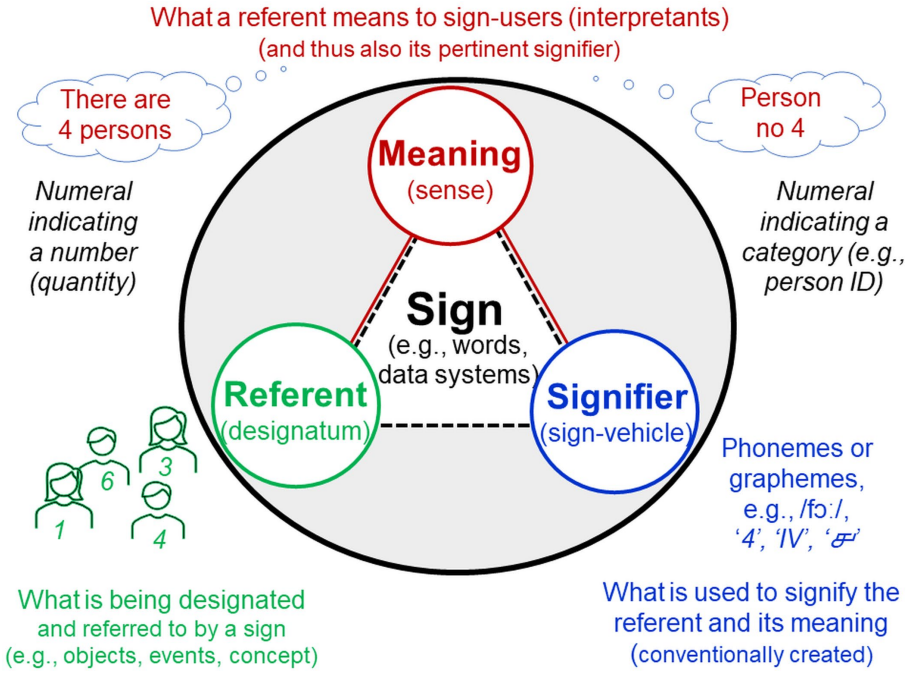


FIGURE 5
The representational function of language and other sign systems. Sign systems comprise three metatheoretical components: signifiers, referents and meanings. Socio-cultural conventions turn publicly accessible (arbitrary) creations (*signifiers*; e.g., phonemes or graphemes) into sign vehicles that can represent objects of consideration (*referents*) and their sense, significance or purpose for someone (*meaning*) even whilst these objects are absent and without any inherent relations to them (e.g., resemblance).

theoretical distinction of study phenomena; Uher, 2013). This is another reason why substantial theories of psychical and behavioural processes (see, e.g., Valsiner, 1997; Sato et al., 2009; van Geert, 2011) are still scarce as some psychologists focussing on constructs and rating-based research noted even themselves (e.g., Haig and Borsboom, 2008; Kyngdon, 2011).

The confusions around constructs are also rooted in the intricacies of language.

Problem complex §8. Naïve use of language-based methods: Reification of abstractions and studying merely linguistic propositions

Language is human's greatest invention (Deutscher, 2006). With words, we can refer to objects of consideration even in their absence and although what we say or write (the *signifiers*) typically bears no inherent relations (e.g., resemblance[14]) to the objects referred (the *referents*). This representational function of words arises from socio-cultural conventions that establish signifier-referent relations, which are merely *conceptual* and therefore not directly apparent but which the sign-using person must learn and know (mentally represent).

Each sign (e.g., word, symbol) thus consists of and involves interrelations between three different components: (a) a signifier (sign-vehicle), (b) a referent (designatum) and (c) the meaning (sense) linking them. Specifically, (a) *signifiers* are physical, publicly accessible phenomena that are often arbitrarily and conventionally created (e.g., graphemes, phonemes) and that we use to refer to particular (b) *referents*, which can be anything that persons can perceive and/or conceive of (e.g., objects, events, ideas, concepts), thus any kind of phenomenon.[15] The (c) *signified* is the meaning that the referents, and thus also the signifiers signifying them, have for the sign-using persons (interpretants)—individually at a given moment but also in socio-linguistic communities over time and contexts—and which is a psychical phenomenon in itself. This psychical component establishes the functional signifier-referent-meaning interrelations[16] on which signs are based and from which new properties emerge that are not present in each of these three components in itself (Axiom 1; Figure 5). This metatheoretical

14 With very few exceptions (e.g., iconicity, onomatopoeia, indexicality).

15 See footnote 9.

16 In semiotics, signifier-referent relations are explored in the field of *semantics*; signifier-meaning relations in *pragmatics*, and interrelations between the signifiers of different signs in *syntactics* (Morris, 1938; Rød, 2004).

concept of signs (building on Ogden and Richards, 1923; Morris, 1938; Dewey, 1946; Peirce, 1958; Vygotsky, 1962; Rød, 2004) illustrates that language involves psychical phenomena *in itself* and is thus inseparable from its users' minds (Axiom 3; Valsiner, 1997). To highlight this, sign systems are called *semiotic representations* in the TPS-Paradigm (Uher, 2015a, 2016b, 2021a).

The representational function of language and other sign systems is essential for abstract thinking because it allows individuals to turn—on mere conceptual levels—perceivable properties (e.g., bitter) into hypothetical objects (e.g., 'bitterness'), thereby making them conceptually independent of their embodied experience (hypostatic abstraction; Peirce, 1958, CP 4.227). These reified (objectified) properties can become objects of consideration in themselves (e.g., 'taste') and can be linked to other perceptions, objects and meanings (e.g., 'bitter' as socio-emotional category). This allows individuals to mentally handle abstract ideas and to abstract them further. More abstract words therefore refer to ideas and concepts that are more distant from immediate perception and that cannot be easily traced anymore to their formerly concrete references and contexts. Hence, words carry meanings that are drawn from their logical connections with other words in the semantic space of a language (which can be depicted in *semantic networks*[17]) as well as from the linguistic contexts in which they are being used (e.g., sentence, paragraph). Put differently, words are basic linguistic units carrying meanings that are drawn from the logical, semantic and meaning-making structures of a language. These structures follow particular rules (language-games; Wittgenstein, 1922), which are shared within socio-linguistic communities to enable communication and which are needed to infer the particular meaning that a person may want to express in language (Peirce, 1958; Vygotsky, 1962; Deutscher, 2006; Neuman et al., 2012; Uher, 2015a, 2016b; Khanam et al., 2019).

Rating methods capitalise on these extraordinary abilities of language. Yet many psychologists know surprisingly little about sign systems. Given this, and mislead by the ease of using language (Axiom 3), they often overlook the inherently representational and composite nature of signs—a classic example of competence without comprehension (Dennett, 2012; see Arnulf, 2020). Perhaps therefore, a sign's most directly apparent component, its signifier (e.g., what is written) is often equated with the entire sign (even if just implicitly). This entails various fallacies, such as when signifiers (e.g., printed item wording) are assumed[18] to carry *in themselves* the meaning that can be ascribed to them (*signifier–meaning conflation*; Table 1) as reflected in the common idea that standardising rating items could standardise also their meanings for raters (as needed for quantification; Problem complex §10 Quantificationism). Ignoring a sign's meaning component can also lead to mistake the signifier for its referent, such as when rating items are equated with the behaviours they describe (*signifier–referent conflation*; Table 1) as done in operationalism (Problem complex §6) and leading to constructification (Problem complex §7). But without conceptual signifier–referent–meaning interrelations, a signifier (from Latin *signum* for mark, token)—literally—cannot signify anything (Uher, 2016b, 2018a, 2021c,d). Language-based methods, such as rating 'scales', are inherently interpretive and context-sensitive, involving individual and changeable meaning construction (Valsiner et al., 2005; Rosenbaum and Valsiner, 2011). This must be considered when aiming to explore the individual experiences that persons aim to express through language (Stulík, 2022).

Originally, rating 'scales' were conceived as capturing just verbal behaviours, whereby verbal declarations were taken as socially accepted symbols for overt acts (Likert, 1932)—an idea refuted almost contemporaneously (LaPiere, 1934). Ultimately, every study phenomenon can be verbally described—otherwise it cannot be researched (Wittgenstein, 1922)—and many psychical phenomena are accessible only through language. Rating 'scales' are often treated as if they could capture just any research phenomenon as long as it is describable in small chunks of colloquial language, reflecting an "anything goes" research attitude. Mere hand-movements for ticking boxes (Baumeister et al., 2007) are thereby conflated with raters' semantically guided meaning construction, beliefs and intuitive judgements encoded in everyday language, which can lead only to pseudo-empirical findings (Smedslund, 1991, 2016). Indeed, in studies using language processing algorithms, more than 86% of the statistical variation obtained in Likert 'scale' responses was *a priori* predictable from the items' semantic fields of meaning (Arnulf et al., 2014).

Availability of a word leads to assume that its referent constitutes a concrete entity (Problem complex §1f Psychologists' own role in their research). This may be true for words denoting concrete referents that are directly perceivable without reflection but not for "fictitious" words such as those denoting abstractions (Jeremy Bentham, 1748–1832, cited in Ogden, 1932). Linguistic abstractions, such as single word terms for constructs (e.g., 'openness'), are often mistaken for real concrete entities—the *fallacy of misplaced concreteness* (Whitehead, 1929). This

17 Semantic networks are a logic-based formalism to describe structures in the mental organisation of knowledge representations and information retrieval pathways (Pirnay-Dummer et al., 2012), whereby the mental association strength between concepts (studied, e.g., with reaction times) is illustrated in nodes and directed edges. These networks describe multidimensional associations between constructs and their nested conceptual structures—thus, the fields of meaning that are common in particular sociolinguistic communities.

18 Reflected in some psychometricians' statements, such as "what really matters in validity is how the test works, and this is [...] a property [...] of the measurement instrument itself (i.e., *of the concrete, physical thing that you can drop on your feet, rather than of a linguistic entity*)." (Borsboom et al., 2009, p. 149, italics added).

linguistic *reification* promotes the conflation of the study phenomena with the constructs used to explore them (construct–referent conflation; Problem complex §7 Constructification) and may mislead researchers to take descriptions of the study phenomena for their explanation (Table 1), resulting in *explanatory circularity* (Mischel and Shoda, 1994; Uher, 2013). Without considering the complex role and function of language in science and everyday life (Axiom 3), rating-based research runs the risk of studying merely linguistic propositions (Wittgenstein, 1922)—thus, only laypeople's knowledge *about* the phenomena of interest but not these phenomena in themselves, the fallacy of constructification (Problem complex §7).

Problem complex §9. Variable-based psychology and data-driven approaches: Overlooking the semiotic nature of 'data'

Failed distinction of study phenomena from study means is also reflected in the disparate use of the term 'data' (Table 1). On the one side, psychologists refer to the study phenomena in themselves (located, e.g., in individuals) as 'data' that are to be observed or collected. On the other side, they consider the variables and values carrying information about these phenomena (located, e.g., on spreadsheets) as 'data' to be statistically analysed. Whatever meaning one may prefer, 'data' cannot refer to both without conceptually conflating disparate elements of research (Uher, 2021c). Analogous problems and conflations occur with the term 'variable' (Table 1). The common jargon around variables is intended to achieve a certain formalisation needed for statistical analysis (Problem complex §11 Statisticism). But it implies that either psychologists study only 'variables' but not individuals or that 'variables' somehow exist in individuals or the world as quantitative entities readily available for statistical analysis (Danziger and Dzinas, 1997; Maraun and Gabriel, 2013; Uher, 2021d).

Data are conceived in the TPS-Paradigm as the sign systems that scientists use to semiotically encode (in *signifiers*) information (*meaning*) about their study phenomena (*referents*; Figure 5). As signs, data can be stored, manipulated, decomposed and recomposed, thus analysed in ways not feasible for these phenomena in themselves (e.g., behaviours). That is, data variables and values are sign systems that are explored *in lieu of* the actual study phenomena and the analytical results obtained from these data, in turn, are used to make *inferences back to* these phenomena (Figure 6). Valid inferences from analytical findings presuppose that it is known what information the data variables and values actually represent, and how exactly they represent the

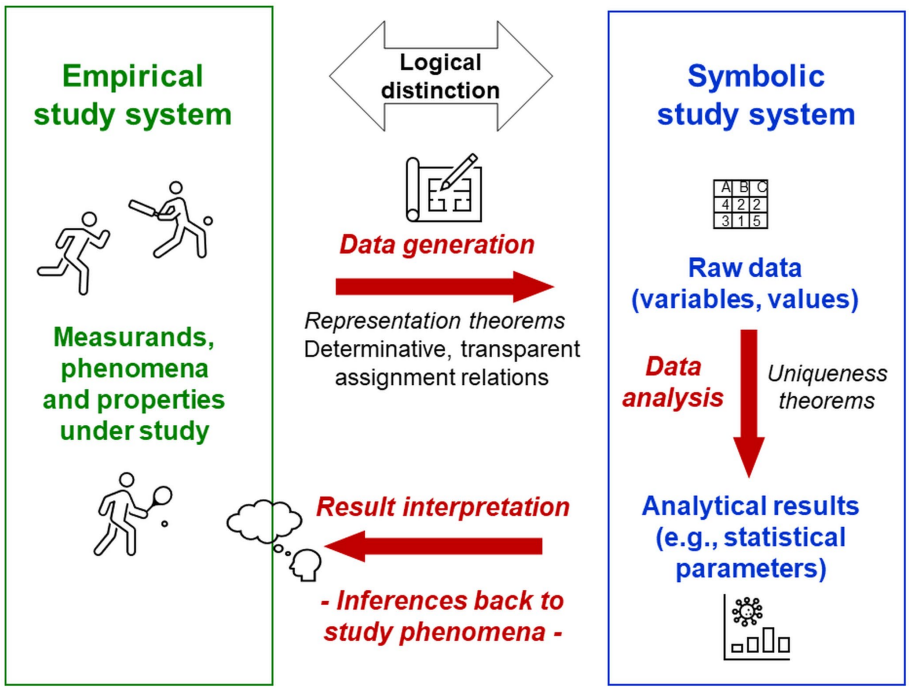


FIGURE 6
Data generation, data analysis and result interpretation. The principles of *data generation traceability* and *numerical traceability* specify rationales by which transparency can be established in the relations between the empirical study system (e.g., persons' behaviours) and the symbolic study system (e.g., raw data variables, values). Still needed is the development of analogous principles of *data analysis traceability* specifying rationales for the transformations that are made within the symbolic study system through different kinds of analytical methods. Such general principles will help establish transparency in the analytical results' relations to the original raw data with regard to the information that these reflect about the measurands and their qualitative and quantitative meanings (e.g., rationales for grouping cases, choosing units of aggregation), thereby guiding researchers' result interpretation with a clear focus on the empirical study system (Problem complexes §11 Statisticism and §12 Nomotheticism).

phenomena and properties studied—thus, *transparency in data generation* (§10 Quantificationism; Uher, 2018a, 2021d,c, 2022). This may appear trivial and obvious. But many psychologists' naïve use of language-based methods (Problem complex §8) and mistaken dualistic concepts (Problem complex §3) lead them to overlook serious problems with their data (Axiom 3).

Data variables that represent information about constructs (e.g., 'extraversion'; 'sex/gender'), given their multiple referents, are called *collective variables* (Thelen and Smith, 1994). They encode information such as can be obtained with inferential rating items but also construct indices created by summarising scores across several item variables (e.g., 'extraversion' scores). Collective variables entail problems that seriously undermine the meaningfulness of statistical analyses and result interpretations. Specifically, from analysing collective variables, it is impossible to differentiate information about the relations of a single construct referent from those of the conglomerate of all referents considered for a construct. This may mask actual causal relations so that statisticians may misleadingly understand the conglomerate of all construct referents as a cause instead of just one single referent, or vice versa. Causality may also be erroneously attributed to one specific referent although another referent of the same construct is actually important. For example, the blended construct 'sex/ gender' may refer to genetic, hormonal, bodily, behavioural and social differences; but which ones are actually relevant and causally related to a study phenomenon cannot be analysed from the corresponding collective variable (for details and further problems, see Toomela, 2008).

Rating 'scales' obscure these problems because items are used *both* as description of the study phenomena and as the (item) variables used to encode and analyse information about these phenomena. This dual use, although seemingly efficient, builds on psychologists' cardinal error, operationalism (Problem complex §6), construct–referent conflation (Problem complex §7 Constructification) and variable thinking (Problem complex §9) and anchors these problem complexes directly into rating 'scales'. Their dual use leaves the specification of a sign system's (1) referents, (2) signifiers and (3) the meanings attributed to and thus linking these two latter to raters' intuitive unknown decisions. This entails serious methodological problems because lack of transparency in *data generation* cannot be remedied by even the most transparent, sophisticated and robust *data analysis* of any preregistered study. Transparent data generation requires specification of (1) the system of the empirical phenomena studied (the referents; e.g., behavioural acts), (2) the symbolic study system used to encode and analyse information about that empirical system (the signifiers; e.g., variable values on spreadsheet), and (3) determinative assignment relations between these two study systems (their meanings), so that the same symbol always encodes the same information about the empirical phenomena (Figure 6; Uher, 2018a, 2021a, 2022).

This idea is basic also to representational theory of measurement, which formalises axiomatic conditions by which empirical relational structures can be mapped to symbolic relational structures (in representation theorems) as well as permissible operations for transforming the latter without breaking their mapping onto the former (in uniqueness theorems; Krantz et al., 1971; Vessonen, 2017). Uniqueness theorems are well-familiar to psychologists (Figure 6), such as for selecting statistical tests that are appropriate for specific data types. But psychologists often overlook (e.g., Borsboom and Mellenbergh, 2004) the fact that explicit representation theorems are basic to any scientific data generation and essential first steps[19] of measurement (Problem complex §10 Quantificationism; Uher, 2018a, 2020b, 2021c,d). Without appropriate consideration of the inherently semiotic nature of 'data' and without explicating their relations to the actual study phenomena, so-called 'data-driven' or 'data-oriented' approaches can only further institutionalise the problems highlighted here.

Problem complex §10. Quantificationism: Numeralisation instead of measurement

Much of psychological theory, research and practice relies on quantitative data—thought to be precise, reliable, logic and objective, enabling rigour, standardisation, clear communication and mathematical analysis. The introduction of quantitative approaches was considered an essential means to emulate the physical sciences' successes and establish psychology as a natural science. Behaviourism and the large-scale assessment industry promoted quantitative approaches as a way of making analyses and decisions independent from the judgement of single experts (Axiom 3). Responsibility for analytical work now lay with instruments, techniques and mathematical-statistical models as unprejudiced tools available for public scrutiny. These scientific methods seemed to enable objective explorations of psychical phenomena in which interpretation and subjectivity hardly played a role anymore (Maslow, 1946; Brower, 1949; Strauch, 1976; Haggerty and Porter, 1997; Westerman, 2006).

These promising prospects drew psychologists' attention to these tools' technical correctness, away from questions about their appropriateness and relevance for the study phenomena (Maslow, 1946) and thus also from elaborating the philosophical and theoretical foundations of these tools—that is, their underlying methodology. Quantification became viewed as a positive value *per se* and a quantitative answer as *generally* better than a qualitative one—a belief known as *quantificationism* (Strauch, 1976). Accordingly, psychologists devised quantitative methods

19 Representational measurement theory, however, provides no concepts and procedures for implementing such theorems and does not consider important elements of measurement (e.g., error and accuracy) and is therefore, insufficient for measurement in itself (Borsboom and Scholten, 2008; Mari et al., 2017).

that were feasible in their field and that they considered to be analogous to physical measurement—yet without checking if these adaptations actually met (1) the criteria of measurement and (2) the peculiarities of their study phenomena (Strauch, 1976; Valsiner, 2017b; Uher, 2018a, 2020b). Specifically, when operationalists defined a study phenomenon's meaning primarily by the operational procedures enabling its investigation (Problem complex §6 Operationalism), application of quantitative methods implied the invalid *a priori* answer that "Regardless of what it is, it can be measured—it is a continuous quantity" (Hibberd, 2019, p. 46; Strauch, 1976; Michell, 1997). Operational procedures yielding convergent numerical results were now interpreted as valid instruments for "measuring constructs." This required (1) the creation of several similar, thus redundant operations for generating quantitative data for the same construct (because, in organisms able to memorise and learn, Axiom 1, possibilities for controlled identical repetitions are limited), (2) methods for analysing the results' empirical convergence, and (3) rationales for interpreting the found convergences' meaningfulness.

The manifold nuances of semantics made rating methods ideal to design at libitum[20] similar and redundant operations to study constructs empirically (Problem complex §6 Operationalism). Aiming to emulate measurement scales, psychologist created verbal 'scales' featuring multi-stage answer categories (e.g., 'rarely, 'sometimes, 'often') that are rigidly scored as numerical values (e.g., '1', '2', '3'). Stevens' (1946) definition of four categories of variables (e.g., nominal, ordinal, interval), indicating information of different levels of complexity (e.g., categorical or sequence information without or with equal intervals), justified the attribution of quantitative properties to such 'scales' and the scores obtained with them. To identify sets of item 'scales' yielding maximal convergence between their operational results and to analyse the obtained scores' consistency (reliability), psychologists devised numerous methods of statistical analysis. These mathematical-statistical procedures, although diverse, are largely uncontroversial—much in contrast to the rationales to determine the scores' meaningfulness (validity) as 'measurements' of a construct. Psychologists still debate whether validation involves concurrent or predictive convergence with scores obtained for theoretically related constructs (Cronbach and Meehl, 1955), or rather the scores' causal relations with the study phenomena (Borsboom, 2005), their plausibility, coherence and appropriateness (Kane, 2013), or the social and ethical consequences of their use (Messick, 1995) and whether validity actually refers to the scores or rather to the instruments used to generate them (Newton, 2012; Kane, 2013).

The basic rationale of quantitative approaches in psychology appears to be coherent at first sight. But it builds on a dense network of misconceptions and conflations that support each other, codified in psychological jargon and woven together through operationalism (Problem complex §6), constructification (Problem complex §7), naïve use of language-based methods (Problem complex §8) and variable thinking (Problem complex §9). This makes it difficult for researchers using these approaches to become aware of the underlying problems and to break out of the intuitive conceptual back-and-forth switching that masks the logical gaps between the different research elements conflated (Axiom 3; Figure 2). These problems therefore emerge time and again in psychology's controversies and crises—such as in those over psychological 'measurement'.

What is measurement actually? Basic criteria across the sciences

When psychologists use operational indicators to 'measure' a construct and interpret the results as reflecting quantifications of 'it', they clearly see the construct as their actual study phenomenon—overlooking that scientific constructs are just means of exploration (Problem complex §7 Constructification), thereby committing psychologists' cardinal error and the logical errors of operationalism (Problem complex §6). The common idea of "measuring constructs" also reflects profound misconceptions about measurement. Psychological 'measurement' is often thought to require "the assignment of numerals to objects or events according to some rule" (Stevens, 1946, p. 667)—an idea easily implemented with rating 'scales'. But this oversimplified definition ignores the basic ideas of measurement, which—across all sciences—are reflected in common interpretations of results that are thought to be obtained through 'measurement' (whether or not the specific procedures used actually justify these interpretations). These ideas were therefore used in the TPS-Paradigm to formulate two abstract, general criteria as *basic common denominators that characterise, across sciences, a data generation process as measurement*. These criteria are:

(1) *Justified attribution of the results to the measurands* (i.e., the specific entities to be measured) and not of something else (as well)—the ontological claim; and
(2) *Public interpretability of the results' quantitative meaning* with regard to the property measured—the semiotic claim.

These criteria underlie two different yet interrelated methodological principles[21] for establishing data generation

20 By comparison, operations suited for observations of behaviours are far more constraint as they are bound to individuals' bodies and the moments and situations in which they occur. Together with the strong fluctuations of behavioural phenomena, this requires intense observer training and makes observations costly and labour intense.

21 These principles, elaborated in the TPS-Paradigm on the abstract methodological level, underlie the frameworks of physical measurement in metrology. Data generation traceability is not explicitly formulated in metrology but implied by the process structures established. Numerical traceability is the methodological principle underlying metrological traceability, which is foundational to the International System of Units (Système International d'Unités, SI). This system codifies the internationally established measurement units and their reference quantities together

processes that enable measurement and for distinguishing these from other processes of evaluation (e.g., opinion making, assessment). These are the principles of data generation traceability and numerical traceability[22] (Uher, 2020b, 2022).

Data generation traceability: Establishing causal measurand—result connections

The term '*measurand*', although fundamental in measurement terminology (JCGM200:2012, 2012), is unknown to most psychologists. Constructification (Problem complex §7), operationalism (Problem complex §6), inferential rating items (Problem complex §8 Naïve use of language-based methods) and variable thinking (Problem complex §9 Variable-based psychology and data-driven approaches) shifted out of focus concrete study phenomena and with them the necessity to specify the concrete entity to be measured—the *measurand*. But what is it that psychologists actually aim to 'measure'?

Other than psychological jargon implies (e.g., "measuring behaviour[23]"), objects or phenomena cannot be measured in themselves; only some of their properties can be. Objects and phenomena often feature various properties (e.g., individuals' bodies feature the properties of length, mass, temperature, etc.; walking behaviour features various temporal and spatial properties, etc.). Therefore, researchers must specify which particular property they study—the *target property*. Of a given target property, in turn, any given study object or phenomenon can feature several entities. One cannot "measure an individual's length" *per se* but only the length of its body height, left forearm, stride, step, distance walked per hour, etc. Hence, scientists must specify which particular entity of the target property in the given study object or phenomenon is their measurand (Figure 7). Psychologists, by contrast, often specify just the researchee as the entity to be studied (e.g., "measure an individual's level of 'activity'").

To justify the attribution of results to the measurands, the entire data generation process must be fully transparent and traceable. This requires operational structures (methods), often implemented through measuring *instruments*,[24] that enable an empirical interaction with the measurand and that establish *proportional, thus quantitative relations* between the measurand

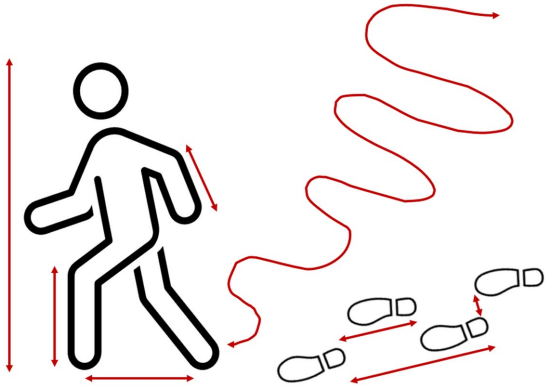


FIGURE 7
Measurands in individuals. Measurands are the specific entities to be measured; their unknown quantity is to be determined through measurement. Scientists must specify which particular entity of a target property (e.g., length) is the measurand in a study. For example, one cannot "measure an individual's length" *per se* but only the specific entities of length that it may feature, such as the length of its body height, left lower leg, left forearm, stride, step, distance walked per hour, etc. In rating-based research, by contrast, psychologists often specify just the researchee in general as the entity to be studied (e.g., "measure an individual's level of 'activity'")—a reflection of the inherently conceptual level of consideration taken.

and the result assigned to it. For measurands that are accessible only indirectly, such *causal measurand–result connections* are established through sequential empirical interactions between different properties forming a connection chain, whereby the result of each interaction step depends on that of the preceding step (e.g., indirect measurement). For example, measuring an object's weight with a spring scale involves the connection chain mass >>> gravity force >>> length of spring deflection (each connected through physical laws) >>> length of extension over the measurement scale (connected through visual comparison) >>> numerical values that measurement-executing persons assign in relation to that scale (connected through semiotic encoding). Unbroken documented connection chains allow a result to be traced, in the inverse direction, back to the measurand, thereby justifying the attribution of this result to that measurand (Figure 8; Uher, 2018a, 2020b).

But what specific values are to be assigned and why?

Numerical traceability: Establishing known quantity—result connections

In measurement, numerical values are used to convey in publicly interpretable ways information on the specific quantity determined for a measurand. To establish this semiotic function (see Figure 5), scientists conventionally define for each given target property (e.g., length) (1) particular quantity references (*referents*; e.g., the specific length of 1 metre), which also serve as measurement units (e.g., metre, yard, mile); (2) specific numerical values (e.g., '1', '1.09361') that are used to indicate defined

with conventional definitions of their interrelations (BIPM, 2006; Czichos, 2011; Uher, 2022).

22 Previously called the set-theoretic and the algebraic requirements of measurement (e.g., in Uher, 2015c,d).

23 Behavioural researchers (see footnote 1) know that behaviour cannot be measured in themselves—only specific behavioural acts shown by a specific individual at a particular time and place can be measured, for example, regarding their temporal (e.g., duration, frequency) or spatial properties (e.g., distance).

24 With increasing knowledge, this also requires theoretical explanations of the interactions enabled by these operations, whereas initial stages of instrument development often rely on mere instrumentalist approaches.

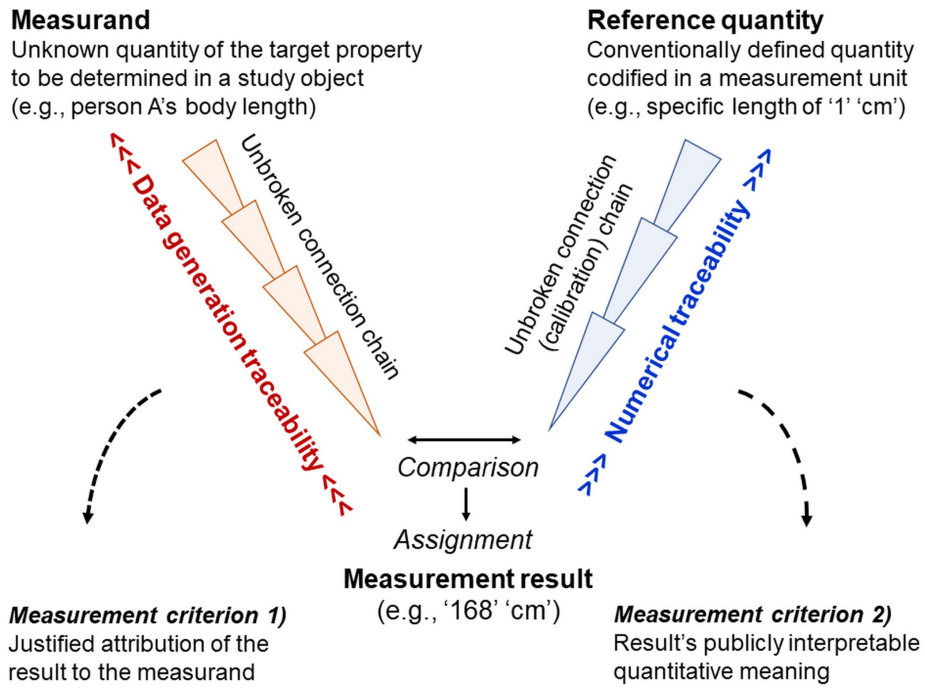


FIGURE 8
Data generation traceability and numerical traceability. The two methodological principles of data generation traceability and numerical traceability and their relations to the two basic criteria of measurement that are (implicitly) used across sciences.

quantities of that target property; as well as (3) empirical interrelations between different quantity references (and between their respective units) codified in non-contradictory mathematical equations (e.g., '1' 'metre' = '1.09361' 'yard').

To ensure the quantitative *meaning* of these numerical values across time and contexts, primary quantity references (e.g., international prototype metre) are defined and connected with all pertinent working references used for measurement execution (e.g., all yardsticks) through networks of unbroken documented connection (calibration[25]) chains (BIPM, 2006). Implemented in measuring instruments, the reference's known quantity is set in proportional relation with the measurand's still unknown quantity, therewith allowing the determination of the latter (Figure 8). Quantity references can be defined arbitrarily[26] and, albeit in different ways, also in psychology, such as counts of test responses of defined correctness (e.g., in attention or achievement tests; Uher, 2020b, 2022).

These two traceability principles—on their methodological (that is, philosophical and theoretical) level of consideration—guide the necessary methodical (that is, operational and procedural) adaptations to the peculiarities of different disciplines' study phenomena (considering also further key elements of measurement not discussed here, such as error and uncertainty; see, e.g., Giordani and Mari, 2014). Their implementation in necessarily discipline-specific theories and practices allows measurement results to be traced back to (1) the measurands (data generation traceability) and to (2) reference quantities, which specify the results' quantitative meaning regarding the target property in publicly interpretable ways (numerical traceability; Figure 8). This allows researchers to make the entire measurement process *and* the results transparent and reproducible (Uher, 2020b, 2021c,d, 2022).

These principles are now used to scrutinise numerical data generation with ratings.

Rating 'scales': Numeralisation instead of measurement

Inferential items do not specify concrete phenomena and properties to be judged (Problem complex §8 Naïve use of language-based methods), but some answer 'scales' indicate a target property, such as frequency 'scales' or the popular agreement (Likert) 'scales'. But can agreement reasonably be assumed to be a

25 These are called calibration chains because, along the connections in the chain, they specify uncertainties as a quantitative indication of the measurement quality of a result to assess its reliability (JCGM100:2008, 2008).

26 The history of measurement in physics and metrology shows that reference quantities can be defined arbitrarily, while still allowing the establishment of a coherent system through systematic unbroken connection networks. The knowledge gained from decreed measurement units and the advent of modern technologies ultimately allowed scientists to replace arbitrary definitions with artefact-free definitions that are based on natural constants and thus reproducible any time and place (e.g., meter by speed of light; BIPM, 2006; Quinn, 2010).

property of phenomena as diverse as those described in constructs of 'anxiety', 'honesty' or 'extraversion'? In judgements of physical properties, such as of the length of lines in Asch's (1955) classic experiment on social conformity, it is obvious that agreement is not a property of these lines but of the persons judging them (Axiom 3). But disentangling the mental abilities that are inherently involved in such judgement processes (e.g., working memory, mental complexity, time perception, self-knowledge) from the specific psychical phenomena to be judged (e.g., beliefs, feelings)—thus, the study means from the study phenomena—is difficult. Indeed, to what extend well-documented individual differences in these mental abilities (John and Robins, 2022) influence raters' abilities to provide graded judgement has hardly been studied (Toomela, 2003). No wonder that, in rating-based research, measurands are commonly left unspecified. Often, the entity to be judged is specified only as the researchee *in general* (e.g., how someone *is*). This inevitably requires abstractions from the momentary occurrences of (psychical or behavioural) phenomena; thus, their consideration at construct level (Problem complex §10 Quantificationism; Whitehead, 1929; Kelly, 1955).

Raters must form and indicate their overall judgements using a bounded set of (mostly) verbal answer categories indicating staged degrees of the enquired property (e.g., frequency) in general, abstract words (e.g., 'seldom', 'sometimes', 'often'). But *how often* is 'often' for a behaviour to occur given that occurrence rates generally vary between behaviours and situations (e.g., chatting vs. laughing in a café vs. hospital; Uher, 2015b)? Regardless of the *different* phenomena that raters consider for an item and that researchers enquire in *different* items, raters must always fit their judgements into the *same* set of answer categories. That is, raters must assign a broad range of quantitative information *flexibly* to a fixed narrow range of values (e.g., five)—thus, *adapt their judgements to the 'scale' rather than to the phenomena and properties to be judged* (Problem complex §1 Psychologists' own role in their research). This is possible only by constructing for the *same* 'scale' category *different* quantitative meanings, which can distort and even inverse quantitative relations (e.g., chatting 'sometimes' may actually refer to higher frequencies than laughing 'often'; Uher et al., 2013b; Uher, 2015b). This fundamentally contradicts the idea of measurement. For accurate and reliable determination of quantities, physical measurement scale units have unchangeable quantitative meanings and pertinent values can be assigned to measurands without upper limits[27] (Uher, 2022).

Regardless of raters' flexible assignments, psychologists score the verbal 'scales' by rigidly recoding[28] the same answer category (e.g., 'agree') always into the same numeral (e.g., '4'). *Numerals* are *signifiers* (Problem complex §8 Naïve use of language-based methods), that is, sign-vehicles (e.g., graphemes, such as '4', 'IV', '𑁤'[29]) that can signify *numbers*—that is, arithmetic values indicating quantities—but also just categorical information (e.g., phone or ID "numbers"[30]) or even letters (e.g., Roman numerals; Figure 5). Psychologists actually do not assign numerical values *in relation to* a scale's units as in measurement. Instead, they *replace the answer units in themselves with numerals*, thereby creating rating 'scores' devoid of information regarding both the specific property studied (e.g., '4' *of what*—agreement, frequency, duration or intensity?) and the specific quantity of that property that these numerals are meant to indicate (e.g., *how much* of that is '4'?).

A numerical score has no quantitative meaning in itself; a measurement value does—it derives from the measurement unit *in relation to which* the value is assigned (e.g., '4' 'metre' ≠ '4' 'yards' ≠ '4' 'gram') and through which it is conventionally linked with a specific defined quantity of the given target property (numerical traceability). Hence, in measurement, scientists assign not numbers as many psychologists believe (misreading[31] even Steven's simplistic definition of 'measurement') but numerals, which are defined as quantity values of a particular property (Mari et al., 2015; Uher, 2021c,d). Consequently, there is just one correct numerical value to conventionally indicate a specific quantity—otherwise accuracy and precision[32] could not be achieved. Rating 'scales', by contrast, enable only *numeralisation*—the creation of numerical scores without specified referents—neither any measurands, nor defined reference quantities nor even the property under study (Uher, 2022). Instead, which specific numerals are assigned to verbal answer 'scales' depends solely on researchers' study-specific decisions about the structural data format (e.g., number of answer categories, unipolar or bipolar scoring; Schwarz et al., 1991; Simms et al., 2019). But these decisions have nothing to do with the quantities to be determined.

27 Bounded value ranges are also used for some measurement units; but these are either repeating, and thus unlimited (e.g., clock time), or inherent to the target property (e.g., degrees in a circle). Other quantitative categories with bounded value ranges refer to specified samples of unlimited size (e.g., percentages, counted fractions), thus indicating quantity values that are traceable (Uher, 2022).

28 Presenting raters with numerical rather than verbal answer categories does not solve the problems discussed here but only shifts the execution of this recoding to implicit, unknown considerations by raters.

29 These are Arabic, Roman and Tamil numerals all indicating 'four'. Tally marks (e.g., 'IIII') are also signifiers.

30 Numeral–number conflation is also widespread in everyday life.

31 Interestingly, Stevens' definition of 'measurement' correctly refers to the assignment of numerals, not numbers (Stevens, 1946, p. 667).

32 In metrology, *measurement accuracy* denotes how closely the determined value agrees with the measurand's true quantity value (e.g., someone's true body height); *precision* denotes closeness between values obtained in replicate measurements (e.g., body height determined using the same or different yardsticks). *Measurement error* is the difference between measured and true value; *uncertainty* describes the dispersion of values attributed to a measurand, thus the reliability by which the measured value can be attributed to the true value (JCGM200:2012, 2012).

Indeed, psychologists rigidly recode 'scale' categories *in the same ways for all items* of a questionnaire regardless of the phenomena described—another instance where they substitute the study phenomena with knowledge unrelated to these phenomena (Problem complex §1d Psychologists' own role in their research).

Many psychologists seem to be unaware of these problems, believing numerals always indicate numbers (*numeral–number conflation*; Table 1; see similarly *numerology*[33])—and thus, quantities. But what is quantity actually? How many psychologists can clearly define it? Numeralisation misleads many to believe that quantities could exist and thus be treated independently of their qualities, as reflected in the common yet erroneous polarisation of 'qualitative' versus 'quantitative' data, methods, etc. This may also explain why psychologists commonly interpret findings from agreement 'scales' not as reflecting raters' levels of agreement, as encoded during data generation, but instead as quantifications of the diverse phenomena *in themselves* that are described in the items (Uher, 2022). This corresponds to re-interpreting a measurement of, let us say, temperature *ad libitum* as one of length, mass or time. Any quantity is always *of something* (Kaplan, 1964). Study phenomena and properties can be identified as such only by their particular qualities; therefore, all quantitative research ultimately has a qualitative grounding (Campbell, 1974). Quantities (from Latin *quantus* for how much, how many) are divisible properties of entities of the same kind—thus, of the same quality[34] (Latin *qualis* for of what sort; Hartmann, 1964). Entities of equal (homogeneous) quality can be compared with one another in their divisible properties (quantities) in terms of their order, distance, ratio and further relations as specified in the axioms of quantity (e.g., equality, ordering, additivity; Hölder, 1901[35]; Barrett, 2003).

What divisible properties (that is, quantities) could agreement 'scales' reflect? 'Strongly agree' ('5') may certainly indicate more agreement than 'agree' ('4'). But could 'agree' ('4') reflect more agreement than 'neither agree nor disagree' ('3'), often chosen to indicate 'inapplicable' (Uher, 2018a)? Does 'agree' ('4') really reflect more agreement than 'disagree' ('2')? Or are agreeing and disagreeing with something not fundamentally different ideas? Semantically, different qualities can be easily merged into one conceptual dimension (e.g., semantic differentials; Snider and Osgood, 1969; Problem complexes §8 Naïve use of language-based methods and §9 Variable-based psychology and data-driven approaches). But what divisible properties could we identify in abstract, qualitatively heterogeneous (i.e., blended) concepts (Problem complex §7 Constructification)? When aggregating rating scores across items, could answering *1 x 'strongly disagree'* ('1') *and 1 x 'strongly agree'* ('5'), thus, having a split opinion or inversed item interpretation, really indicate (roughly) the *same* level of agreement (averaging '3') like answering *2 x 'neither disagree nor agree'* ('3'), thus having 'no opinion'? The logico-semantic meanings of verbal answer categories—even if just ordinally conceived—are clearly discordant with the quantitative meanings that are commonly ascribed to the numerical scores into which they are recoded (Uher, 2022). Indeed, raters' reasons for ticking off answer boxes are often not quantitative at all but rather trivial and thus different from researchers' 'scale' interpretations (Problem complex §1a Researchers' own role in their research; Uher, 2018a).

The ease of applying Steven's 'scale' categories to rating 'scales' led psychologists to overlook that just (1) specifying a *structural data format* (e.g., five values) and (2) assigning to these values a particular *conceptual data format* (e.g., ordinality) neither enables the (3) necessary traceable *empirical interaction* with the measurand (data generation traceability) nor does it (4) specify a *conventionally agreed reference quantity* determining the assigned values' quantitative meaning (numerical traceability). Measurement scales,[36] by contrast, must fulfil all these four methodological functions, which are needed at different stages of a measurement process and therefore not interchangeable (for details and the disparate meanings and uses of 'scales' and 'units'; see Uher, 2022).

These practices entail that, with rating 'scales', causal measurand–result connections (data generation traceability) cannot be established, precluding the results' attribution to the measurands (e.g., the researchees). Without traceable connections to known quantities (numerical traceability), rating scores have no publicly interpretable quantitative meaning either. The only option to create meaning for such scores is their between-case comparison. This is why statistics became essential to implement quantitative approaches in psychology.

Problem complex §11. Statisticism: Result-based data generation, methodomorphism and pragmatic quantification instead of measurement

Statistical methods enabling between-case analyses quickly became increasingly complex and a discipline of its own. Indeed, for many psychologists, statistical analyses seem to have become more of an end in itself than a means for analysing data obtained to explore questions and to solve problems (Brower, 1949; Lamiell, 2019). This has led to

> … the syndrome that I have come to call *statisticism*: the notion that computing is synonymous with doing research,

33 Johnson (1943) similarly highlighted that numerals [he wrote 'numbers'] are often treated without regard to either their functions (e.g., nominal, ordinal, cardinal) or their characteristics of abstractness or concreteness, which he referred to as *numerology*.

34 Therefore, so-called 'quantitative' and 'qualitative' methods or data cannot be complementary to each other (Axiom 2), as recently suggested (e.g., Ramlo, 2021).

35 For an English translation, see Michell and Ernst (1996).

36 For this reason, 'scales' that to not fulfil all four methodological functions are put in quotation marks in the TPS-Paradigm.

> the naïve faith that statistics is a complete or sufficient basis for scientific methodology, the superstition that statistical formulas exist for evaluating such things as the relative merits of different substantive theories or the 'importance' of the causes of a 'dependent variable'; and the delusion that decomposing the covariations of some arbitrary and haphazardly assembled collection of variables can somehow justify not only a 'causal model' but also, praise a mark, a measurement model'. (Duncan, 1984, p. 226; italics added; similarly, Lamiell, 2003, 2013).

Statistical methods build on particular theoretical assumptions. Their contributions to the results cannot be separated from those of the phenomena analysed. Statistical theories may therefore impose structures onto the data that, if erroneously attributed to the study phenomena, may influence and limit the concepts and theories that psychologists develop about these phenomena (Axiom 3). This *methodomorphism* (Danziger, 1985a) is a further instance where study phenomena are substituted with knowledge unrelated to them (Problem complex §1d Psychologists' own role in their research), thereby conflating study phenomena with study means—psychologists' cardinal error.

Specifically, psychometricians develop rating 'scales' (amongst others) enabling the generation of numerical scores that differentiate well (e.g., item discrimination) and consistently between cases (reliability) and in ways considered meaningful (validity). This *result-dependent data generation*, however, aligns both data generation and analytical results to statistical criteria rather than to the study phenomena's properties (Figure 9; Uher et al., 2013a; Uher, 2021d). Common psychometric 'quality' criteria, such as inter-rater and internal reliabilities, concern relations between the generated scores but neither these scores' relations to the measurands (data generation traceability) nor to known reference quantities defining their meaning (numerical traceability). Thus, to create quantitative meaning for rating scores through sample-level statistics, psychologists analyse the relations of scores obtained for *different* individuals, and thus *different* measurands. This basically means comparing scores with unknown quantity information in order to create quantitative

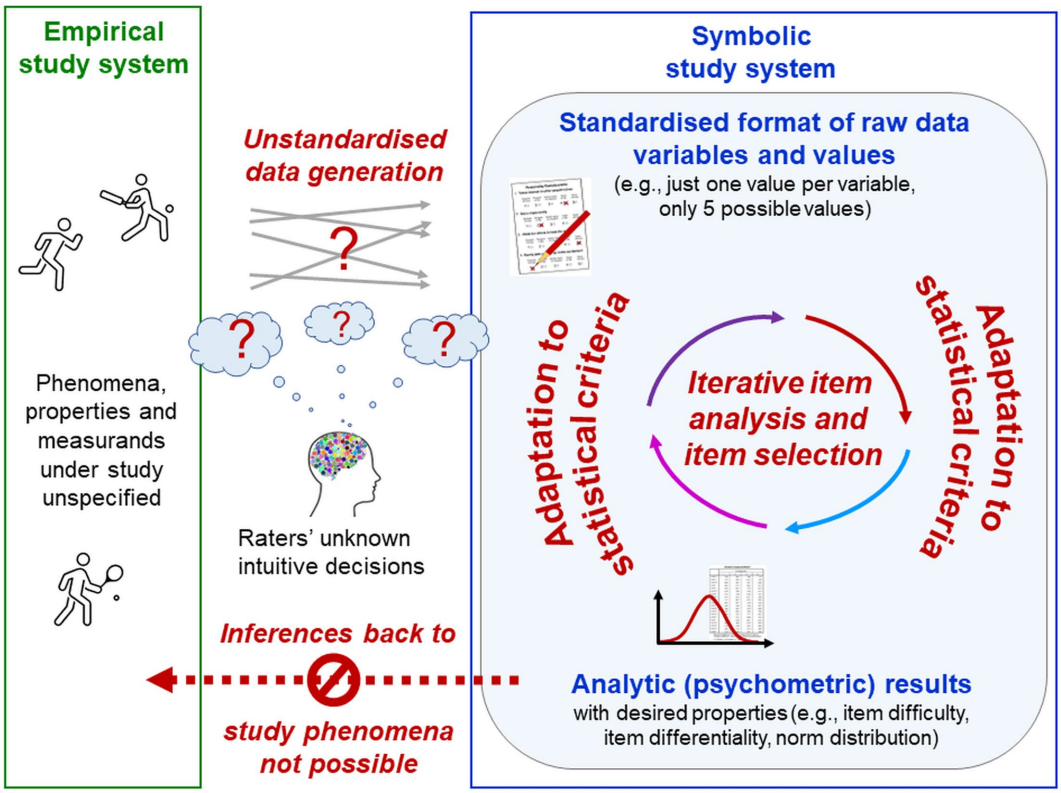


FIGURE 9
Result-based data generation enabling only pragmatic quantification. Psychometric 'scale' development involves iterative processes of item analysis and item selection, in which only those items are retained that allow the generation of data that differentiate well (e.g., item discrimination) and consistently between cases (reliability) and in ways considered meaningful (validity). This *result-dependent data generation*, however, aligns both the data generation and the analytical results to statistical criteria rather than to the study phenomena's properties. This may be useful for pragmatic purposes but it precludes inferences from the analytical results obtained back to the phenomena, properties and measurands under study. These inferences are additionally compromised by the unknown intuitive decisions that raters are making when they generate the data before these are being psychometrically analysed and modelled.

meaning for them—a truly Münchhausenian[37] effort—which therefore fails when all cases score the same. In measurement, by contrast, the measurand's unknown quantity (e.g., an individual's body weight) is compared with that of a known reference quantity (e.g., standard kilogram unit), which establishes the result's conventionally agreed quantitative meaning (e.g., *how* heavy that is; numerical traceability; see Figure 8).

Internal reliability and the validity of rating 'scales' additionally concern relations of scores obtained for *different* items and *different* constructs describing *different* phenomena with *different* properties. Hence, psychological validity theories are about empirical relations of the study phenomena (e.g., those described in an 'extraversion' 'scale') with some *other* phenomena that are considered to be meaningfully related for some reason or purpose (e.g., job performance or phenomena described in another 'extraversion' 'scale'). That is, psychological validity concepts are about relations between phenomena *of different qualities*, whereas measurement is about capturing *quantitative* (divisible) properties of *one specific* quality (e.g., mass; Problem complex §10 Quantificationism). Psychometric scores are useful to discriminate between responses in ways considered to be meaningful but at the expense of unknown relations to the actual measurands in the phenomena studied and of their unknown quantitative meaning. This *utility perspective* is inherent to rating 'scale' development and validation, and appropriate for pragmatic purposes (Dawes et al., 1989; Barrett, 2003). But this *pragmatic quantification* is not measurement and therefore neither justifies the results' attribution to the measurands (e.g., individuals) nor does it establish their quantitative meaning (e.g., *how much* of *what* specifically is it?; Uher, 2021c,d, 2022).

The complex statistics used in quantitative psychology may obscure these problems because every statistical operation removes the analysed data further from the phenomena that they are meant to represent. The more complex the statistics, the more difficult it is to keep track of their connections–which, for rating data, are already seriously compromised (Problem complexes §1–§10)–and thus to check the appropriateness of analyses and interpretations regarding the actual study phenomena (Brower, 1949). Specifically, statistical scores (e.g., effect sizes, correlations) are abstract concepts that describe distributions patterns in a sample and that can therefore inform neither about each measurand's quantity (e.g., single individuals' body weight) nor about the meaning of the quantity determined for a measurand (e.g., *how heavy* that is). Statistics neither is measurement (Fisher, 2009) nor is it therefore needed; indeed, measurement has been successful long before statistics has been developed (Abran et al., 2012).

Psychologists' focus on sample-level statistics also influenced their understanding of nomothetic approaches, which are needed to generalise knowledge about individuals.

37 Referring to the famous story of Baron Münchhausen who pulled himself and the horse on which he was sitting out of a swamp by his own hair.

Problem complex §12. Nomotheticism: Sociological/ergodic fallacy and primacy of sample-based over case–by–case based nomothetic approaches

Individual differences (differential) research is a field of its own, introduced as the population-level investigation of individuals (*differential psychology*)—together with the field devoted to individual-level investigations (*personality psychology*)—by William Stern who founded both sub-disciplines and laid their methodological foundations (e.g., variable-oriented and individual-oriented approaches,[38] Stern, 1900, 1911; see Lamiell, 2003, 2013). Stern recognised that *inter*-individual variation and *intra*-individual variation are equally important characteristics of psychical phenomena. But the American assessment industry, group-based experiments (Danziger, 1985b) and the necessity to pragmatically create quantitative meaning for otherwise meaningless rating scores (Uher, 2021d, 2022) entailed that differential approaches prevail in psychology's terminology, concepts, practices and scientific standards since the early 20th century (Cronbach, 1957; Richters, 2021).

This shifted psychologists' focus away from analysing psychical processes, necessarily located in the single individual, to analysing distribution patterns in populations (e.g., socio-demographically defined). Results were now presented as aggregate data obtained from many individuals (e.g., group averages) without analysing individual patterns (Danziger, 1985a). But instead of attributing their results to the samples analysed, psychologists continued to interpret them with regard to single individuals, which remained their (unlike sociologists') focus of interest and theoretical unit of analysis. This entails the *sociological fallacy*, which arises from the failed consideration of individual-level characteristics when drawing inferences regarding the causes of group-level variability (Diez Roux, 2002).

This inferential fallacy required axiomatic acceptance of *ergodicity*, a property of stochastic processes and dynamic systems, which presumes isomorphisms between *inter*-individual (synchronic) and *intra*-individual (diachronic) variations. Ergodicity fits all invariant phenomena, which do not undergo change or development, and in which simultaneity and successivity are therefore equal. But ergodicity does not apply to psychologists' study phenomena (Axiom 1) as has been proven by

38 Data matrices of X_i individuals by Y_j variables can be explored from two orthogonal viewpoints. *Variable-oriented approaches* focus on variables and study the distributions of their values across all *i* individuals, thus, exploring populations and patterns of between-individual differences but not single individuals. *Individual-/person-oriented approaches*, by contrast, focus on individuals and study each individual's values on all *j* variables, thus exploring individual configurations of values across different variables that can be illustrated as a profile, which reflects a property of the individual but not of the population (Bergman and Trost, 2006; Uher, 2011).

applying classical mathematical-statistical (ergodic) theorems—first derived in ergodic theory, a branch of mathematics originating in statistical physics, in the 1930s (Molenaar, 2008; Molenaar and Campbell, 2009; Valsiner, 2014). Why did this elude 'quantitative' psychologists, despite their keen interests in implementing mathematical approaches analogous to the physical sciences?

Psychologists' common practice to deduce information about *intra*-individual processes from analyses of *inter*-individual variability builds on the *ergodic fallacy* (van Geert, 2011; Speelman and McGann, 2020). Its institutionalisation is most prominently reflected in the common conflation of personality psychology with differential psychology (Table 1), especially in American psychology, where individual differences are regularly equated with individuality (Lamiell, 2013; Uher, 2018c) and where differential analyses of rating data (using variable-oriented approaches) are used by default to study individual functioning and development. From ergodicity it follows, however, that findings from group comparisons or correlations can be generalised only if (1) each individual obeys the same statistical model (assumption of homogeneity) and if (2) the statistical properties (e.g., factor loadings) are the same at all points in time (assumption of stationarity). But, in psychology, these conditions are rarely met (Molenaar, 2004; Molenaar and Campbell, 2009; Salvatore and Valsiner, 2010).

The assumption of *psychical homogeneity*—that all individuals are the same—is logically necessary for sample–to–individual inferences and pragmatically and methodically convenient (e.g., assuming raters' standardised item interpretation). But it is invalidated already by ordinary everyday experience, not to mention a solid body of empirical and theoretical research (Richters, 2021). Indeed, *psychical heterogeneity* is the core idea of 'personality' constructs! The assumption of psychical homogeneity also contradicts *degeneracy*, the capacity of different (non-isomorphic) structural elements to contribute to or perform the same function—a fundamental design principle underlying all complex biological systems (Axiom 1). Degenerate systems feature both *many–to–one structure–function relations* (degeneracy, e.g., polygenic 'traits') and *one–to–many relations* (pluripotency, e.g., pleiotropic 'genes'; Mason, 2010, 2015). This unifying explanatory principle underlies the psychological concepts of *equifinality* and *multifinality*—individuals' capacities to leverage *different* psychical processes and structures to accomplish the *same* behavioural outcome, and vice versa, respectively (Figure 10; Richters, 2021).

This highlights serious problems in the rationales commonly used to generalise findings in psychology. Specifically, the ergodic fallacy misleads many psychologists to understand nomothetic approaches (from Greek *nomos* for law) only in terms of *neo-Galtonian differential approaches* in which sample-level averages are studied and generalised to the individuals thus-summarised (Figure 11; Danziger, 1985b, 1990). But aggregates are statistical constructions of numerical data, constellations of data points featuring structural patterns that have no inherent meaning or theoretical significance in themselves (Richters, 2021).

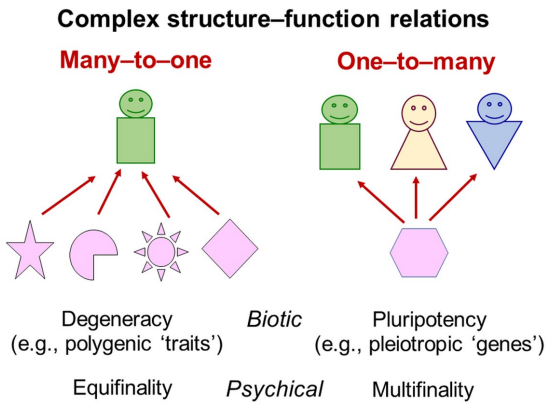


FIGURE 10
Fundamental design principles of complex living systems. Complex living systems feature both many–to–one structure–function relations (degeneracy, e.g., polygenic 'traits') and one–to–many structure–function relations (pluripotency, e.g., pleiotropic 'genes'). This unifying explanatory principle also underlies the psychological concepts of the equifinality and multifinality of psychical phenomena.

Limiting research to *group-level analyses* intrinsically disconnects theory development from descriptions of real individuals and cannot reveal what is common to all. Sample-based nomothetic approaches have turned psychology into a science largely exploring populations rather than individuals—*psycho-demography* (Lamiell, 2018; Smedslund, 2021). They lead to findings and theories that are uninformative about individuals' functioning and development (Danziger, 1990; Lamiell, 2003; Robinson, 2011; Vautier et al., 2014; Smedslund, 2016), an important point increasingly considered also in applied fields, such as decision-making research (Chen et al., 2020).

The often mis-conceived and partly polarised debates about *idiographic* versus *nomothetic strategies of knowledge generation* (Figure 11; Lamiell, 1998; Windelband, 1904/1998), driven by the imperative of establishing psychology as a nomothetic science, mislead many psychologists to overlook that, nomothetic approaches are ultimately based on idiographic approaches—because every science builds on single cases. *Idiographic approaches* (from Greek *ideos* for the peculiar) model local phenomena of single cases in their dynamic contexts. These can then be explored using *case–by–case based nomothetic analyses* to identify generalities that are, indeed, *common to all* cases (Figure 11). Thus, individuals are grouped on the basis of the commonalities and differences that they are shown to exhibit in the study phenomena and properties. Considering degeneracy (many–to–one and one–to–many structure–function relations), the thus-created groups of individuals can then, in turn, be further explored for underlying structures and processes and for commonalities and differences in them. This *Wundtian nomothetic approach*, because it is case–by–case based, allows researchers to develop generalised knowledge and theories about *intra*-individual processes and functioning (Lamiell, 2003; Salvatore and Valsiner, 2010; Robinson, 2011).

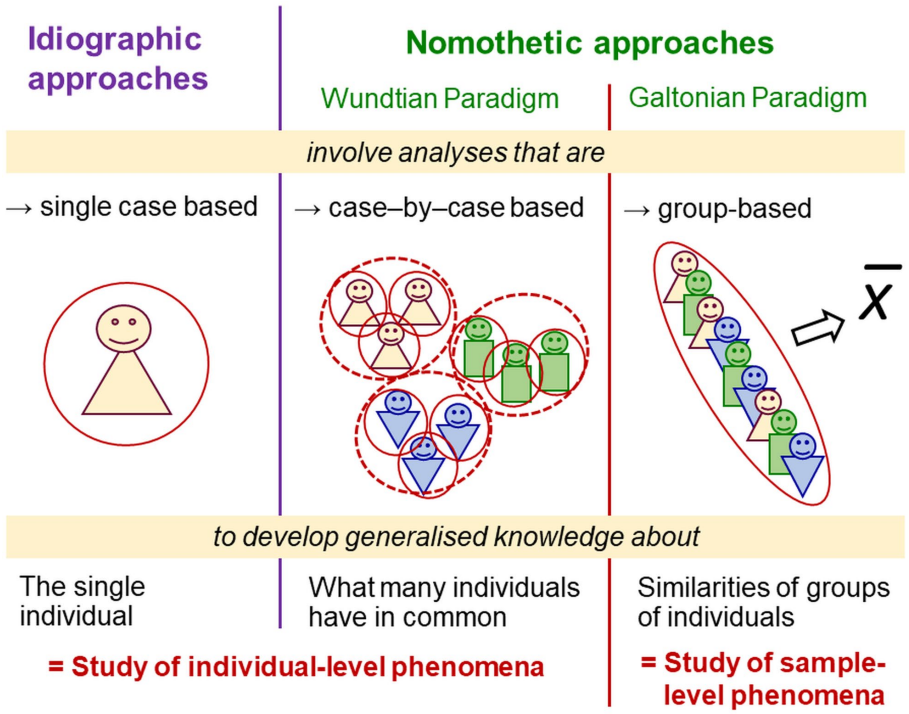


FIGURE 11
Strategies of knowledge generation. Idiographic approaches involve *single case based* analyses. In *case–by–case based nomothetic approaches* (Wundtian Paradigm), individuals are grouped by the commonalities and differences that they exhibit in the study phenomena, which enables the development of generalised knowledge and theories about intra-individual processes and functioning. In *group-based nomothetic approaches* (Galtonian Paradigm), by contrast, individuals are grouped by properties that the researchers find *a-priori* relevant, which inevitably influences and limits the concepts and theories that psychologists can develop about the phenomena studied (methodomorphism) and precludes identifying what individuals have indeed in common.

Galtonian sample-based nomothetic approaches, by contrast, involve analyses in which individuals are grouped by properties that the researchers find *a-priori* relevant (specified as 'independent variables', e.g., 'sex/gender', age), thereby again substituting their own (pre-analytical) knowledge for their study phenomena and assuming these could be understood in terms of categories readily available to the researchers (Problem complex §1d Psychologists' own role in their research). This knowledge inevitably influences and limits the concepts and theories that psychologists develop about the phenomena studied—another instance of methodomorphism (Problem complex §11 Statisticism).

Rating data, given their inherently differential meaning and obscured relations to the study phenomena, contributed to the primacy of sample-based nomothetic approaches in psychology. To explore intra-individual processes, data generation methods able to capture intra-individual variability are needed but still not very common in psychology (Van Geert and Van Dijk, 2002; De Ruiter et al., 2017). With individual-/ person-oriented methods, by contrast, a solid portfolio of analytical methods, grounded in William Stern's methodologies, has already been developed for implementing case–by–case based nomothetic approaches in empirical investigations (e.g., Bergman and Trost, 2006; von Eye and Bogat, 2006).

Conclusions

Measurement is valued in science and society because it provides reliable, accurate and precise information. Measurement in physics is fairly complicated. But with ratings, so it seemed, psychologists have devised a method allowing them to 'measure', with ease and efficiency, almost anything describable in brief colloquial statements through laypeople's judgements. Of course, physicists and psychologists study vastly different objects of research and these necessarily require different methods of research. Certainly, psychology does not need the level of measurement accuracy and precision required for sciences like physics, chemistry and medicine, where errors can lead to airplane crashes, explosions or drug overdoses. And yet, psychologists themselves draw explicit analogies to physical measurement (e.g., in conjoint or Rasch 'measurement'; Trendler, 2019, 2022; Uher, 2021c) and they regard rating 'scales' as 'measuring instruments' with psycho-'metric' 'precision' (Simms et al., 2019) and able to determine judgement 'accuracy' (Kenny, 1991; Funder, 1995).

But does it matter if we call it 'measurement'? It matters because, measurement is not just any activity to produce numerical data. Across all sciences and in society, measurement is regarded a structured documented process that justifies the

results' attribution to the measurands and establishes their publicly interpretable quantitative meaning. These criteria justify the high public trust placed in measurement (Haggerty and Porter, 1997). Psychologists' 'measurement' jargon therefore invites serious jingle fallacies (same term, different concepts) that can mislead researchers, decision makers and laypeople alike. For example, when presented as results of 'measurement', even minor differences are interpreted as meaningful and attributed to the individuals studied (e.g., 'sex/gender' differences; Hyde, 2005). This can have serious consequences, such as when psychometrically determined IQ scores expressed to two-decimal place precision are used in decisions on the death penalty for offenders (Barrett, 2018)—although psychometric scores are mere pragmatic quantifications (Uher, 2021c) that require adjustment to be meaningful (Young et al., 2007; Flynn, 2012). It will only be a matter of time that psychometric scores are challenged in court like forensic psychologists' and psychiatrists' diagnostic practices have been before (Faust, 2012; Barrett, 2018).

Changing the definition of a key scientific activity, such as by 'softening' or 'widening the definition of measurement for psychology (e.g., Finkelstein, 2003; Mari, 2013), cannot establish its comparability across the sciences—but undermines it (Uher, 2020b). Psychologists' 'measurement' jargon and complex statistics gave them a false sense of advancement and of having established a solid scientific framework for their empirical research. Therefore, critical reflection about the meaningfulness and interpretation of the numerical scores produced and of the quantitative analyses applied to them seemed to have become obsolete. Indeed, rating 'scales' are used in virtually identical ways for almost a century now (Thurstone, 1928; Likert, 1932) and differential approaches still prevail. Yet psychology's continued crises about its findings, theories and research practices testify to fundamental problems still unsolved. Current initiatives to tackle these problems (e.g., open and 'meta' science; Malich and Rehmann-Sutter, 2022) are all targeted at improving data analysis and interpretation (e.g., response coding and transformation; construct operationalisation and validity; statistical tests; Flake and Fried, 2020; Hardwicke et al., 2022). But rating 'scales' remained largely unchallenged. They have become even psychology's standard method of data generation (Baumeister et al., 2007) although its philosophical and theoretical foundations—needed to justify this powerful status—have not been elaborated.

The present rigorous analyses of the metatheoretical and methodological foundations on which rating 'scales'—by their very conception, design and application—are built, revealed a dense network of 12 complexes of problematic conceptions, misconceived assumptions, logical errors and faulty jargon (Figure 3). Ambiguous terms with disparate meanings create logical gaps that researchers intuitively bridge through a conceptual back-and-forth switching between the concepts conflated, thereby establishing an internal coherence that makes this network difficult to be identified and recognised by those (unwittingly) relying on it. Through the widespread and uncritical application of rating 'scales', these 12 problem complexes have become institutionalised in a wide range of research practices and therefore cannot be remedied with little quick fixes that many may hope for.

This leaves but one conclusion: Unless ratings are removed from psychology's portfolio of research methods, its recurrent crises (e.g., replication, confidence, validation and generalisability) cannot be tackled. Ratings may be useful for pragmatic purposes in applied fields, but they preclude measurement and—far more importantly—they preclude the scientific exploration of psychology's study phenomena and thus its development as a science.

Recommendations: Directions for solving psychology's crises

As sobering as this account may be, it also opens up directions for development that are needed to solve psychology's crises holistically by getting to the root of the problems rather than just scratching on the surface as many previous proposals. In this final section, specific theoretical concepts, methodologies and methods are derived from each of the 12 problem complexes, guided by the TPS-Paradigm's basic axioms—Complexity, Complementarity and Anthropogenicity—on which the present analyses are based. The solutions proposed and their implications are necessarily overlapping and should therefore be considered holistically across the entire network, although, for a better overview and to avoid redundancies, Table 2 outlines for each problem complex only key implications.

Finally, the author would like to offer some overarching general conclusions about key directions of development in psychology that go beyond the 12 problem complexes; detailed elaborations will be published elsewhere.[39]

Make reflexivity a key qualification of any researcher

No research, including the author's, is free of conceptual errors—simply because errors are human and science is inherently anthropogenic (Axiom 3). Recognising errors in one's own thinking requires self-reflection about one's own positioning as human being in the world and in one's own research work (Bohr, 1937; Danziger, 1985b). Reflexivity should therefore be (re-) established as a basic skill of every researcher (Fahrenberg, 2013; Marsico et al., 2015). This requires reflection also about the embeddedness of science in its societal, political and historical

39 Uher, J. (in preparation). *Overcoming Method-Centrism in Psychology – Beyond Rating Scales*. (Switzerland: Springer Nature).

TABLE 2 The 12 problem complexes, their implementation in rating 'scales' and derivation of directions for possible solutions.

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §1. Psychologists' own role in their research: Unintended influences		
Perspectival disparity between researcher and researchee together with *psychologists' non-independence of their objects of research* (being individuals themselves) entail a confusion of psychologists' own standpoint with that of the psychical phenomena studied. Central *fallacious assumptions:* ■ §1a Intersubjective confusion ■ §1b Attribution of reflectiveness ■ §1c Ignoring study phenomena's relevance in researchee's horizon of their lifeworld ■ §1d Substituting psychical phenomena with knowledge unrelated to hese phenomena ■ §1e Preference of scientific account over that the researchee's ■ §1f Misleading availability of ordinary words.	Raters are assumed to understand and use rating 'scales' similar to the researchers, to consciously reflect on and experience the phenomena as described by the researchers, thus not expressed in their own words and ignoring substantial variations on raters' part. Item contents are chosen to fit researchers' questions, theories and methods (e.g., for 'scale' development); thereby aligning rating 'instruments' to this scientific knowledge rather than to the described phenomena and their relevance for the raters. Everyday language invites all kinds of inferential and attributional biases (Problem complex §8 Naïve use of language-based methods).	**Anthropogenicity (Axiom 3):** Reflect on basic assumptions that may (unintentionally) influence research and that may be (implicitly) attributed to others, especially assumptions encoded in everyday language, which may therefore influence understanding and interpretations of both researchers and researchees. Study the meanings that researchees construct for themselves and as expressed in their own words (e.g., open-ended answers analysed using semantic computer algorithms), rather than using identical wordings and ignoring their inherently flexible and context-depending meanings (Problem complex §8 Naïve use of language-based methods). Ask researchees to reflect on and specify the phenomenon considered, the meaning it has for them in given contexts, and what specifically they considered in their judgements (e.g., using techniques of participant validation).
Problem complex §2. Beliefs in researchers' objectivity: Illusions of scholarly distance		
By virtue of being researchers, psychologists often regard themselves as objectively distanced from the researchees and the study phenomena. But being individuals themselves, they are not independent of their objects of research; their own positioning in the world therefore precludes the possibility of taking a neutral observer standpoint.	Online questionnaires maximally distance researchers from researchees; any contact is just only indirect and virtual. But lack of contact hinders researchers from becoming aware of and exploring possible differences in perspective, presumptions, interpretations (Problem complex §1 Psychologists' own role in their research), thus precluding implementation of any corrective means as is needed to develop concepts of 'objectivity' given the peculiarities of psychical phenomena.	**Anthropogenicity (Axiom 3):** Reflect on and consider the own positioning in the world regarding the study phenomena and the researchees (e.g., using techniques to document reflexivity as in interpretive methods). Conceptualise the social encounter, which every study on individuals constitutes, to explore how researchers and researchees and their roles may influence the results. Study and treat researchees as individual beings, not as anonymous sources of information. Get in direct contact and actually see the researchees in person (just as medical doctors and researchers do).
Problem complex §3. Mistaken dualistic views: Individuals as closed system		
Thought from an observer standpoint, researchers conceive individuals as opposed to and thus separate from the conditions in which these are being studied (e.g., situations, environments). In such *dualistic concepts*, external contexts are categorised by the researchers and ascribed researcher-determined meanings that are assumed to be identical for all researchees, ignoring that their relevance and meaning depend on the given individual.	Common belief that raters would react to item stimuli that have standardised meanings as determined by the researchers and as needed for quantification (Problem complex §10 Quantifiationism) but ignoring substantial and context-dependent variations in raters' understanding and interpretation of the rating 'scales' and the phenomena to which they may refer (Problem complex §1 Psychologists' own role in their research). Raters' responses are flexibly attributed particular meanings, in line with researchers' particular theories, thus involving several of the psychologist's fallacies (Problem complex §1 Psychologists' own role in their research).	**Complexity theories (Axiom 1):** Conceive individuals as complex open systems able to organise themselves in their systemic interdependences with their external contexts. Use *inclusive concepts*, in which relevant parts of an individuals' surrounding and their meanings are identified and conceived in dependence of the researchee's characteristics (e.g., their perception and interpretation of these contexts); such as using *Brunswik's ecological validity concept* (not to be mistaken as similarity of experimental with everyday life situations) to describe the informativeness that elements of the researchees' external contexts have for them. Apply *dialogic and dialectic concepts and theories* to explore individuals' meaning making, functioning and development in their contextual embeddedness.

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §4. Lack of definition and theoretical distinction of study phenomena: Conceptual conflations and intersubjective confusions		
Basic psychological terms and concepts are poorly defined and ambiguous, whilst specific terms and concepts proliferate, causing numerous jingle–jangle fallacies. Lack of differentiation between psychical and behavioural phenomena entails inferential biases, conflation of description and explanation (Table 1), inadequate method choices, and thus the impossibility to explore these phenomena's relations with one another and with other kinds of phenomena.	Rating 'scales' are worded in everyday language, which entails imprecision, high degrees of inferentiality and frequent conflation of description with explanation (Problem complex §8 Naïve use of language-based methods). Many inferential items require raters to judge entities that are generally imperceptible, imperceptible in others, or no longer perceptible, thus forcing raters to rely on memories, beliefs and ideas, thus their everyday constructs (Problem complex §7 Constructification). Which particular phenomena raters actually consider in a rating and thus encode in the data, or if they process the item contents just on the mere conceptual-semantic level remains unknown.	**Complementarity (Axiom 2), Complexity theories (Axiom 1):** Develop frames of reference, methodological approaches and methods that are appropriate for the study phenomena (e.g., accessibility-based conceptual distinction of kinds of phenomena and therefrom derived classes of suitable methods). Ensure that the study phenomena actually occur during data generation and record their temporal, situational and interpersonal contexts, profiting from technical advancements in tracking methods (e.g., reality mining, ambulatory [ecological momentary] monitoring; Problem complex §5 Reductionism). This is essential to explore the complex interplays between different kinds of phenomena and their emergent features in higher-order and more complex phenomena (e.g., actions, language, 'extended mind'), embedded in multi-level contexts.
Problem complex §5. Reductionism: Category mistakes, atomistic fallacy and decontextualisation		
Incorrect reductions are frequently made in psychology (often in conjunction with Problem complex §4 Lack of definition and theoretical distinction of study phenomena): *Ontological reductionism*: The claim that complex phenomena could be described in terms of more fundamental ones. *Epistemological reductionism*: The claim that higher-level phenomena could be explained by lower-level phenomena; this entails category mistakes. *Methodological reductionism*: The claim that complex systems could be understood by dissecting them into their supposedly isolable building blocks (*elementarism*). This *decontextualisation* entails the *atomistic fallacy* (from information obtained at lower levels, incorrect inferences are made at higher levels of organisation). It also entails that the interrelations and functioning of the isolated elements in their systemic contexts cannot be studied. As a result, description is used as and conflated with explanation (Table 1).	Rating data are often erroneously interpreted as reflecting 'psycho-physical mechanisms', often called 'traits', which are assumed to be adaptive, inherited and universal across cultures. Concepts are broken down into brief, colloquially worded, often decontextualised statements, presented in mixed order, which precludes holistic, contextualised considerations and allows only for abbreviated reflection. Rating items radically dissect and decontextualise information about the study phenomena, thus involving the atomistic fallacy and precluding investigations of structural and processual patterns as they occur in the study phenomena. Instead, these chunks of verbal information are put together using statistical procedures, thus using knowledge unrelated to the study phenomena (Problem complexes §1 Psychologists' own role in their research; §10 Quantificationism and §11 Statisticism) and also unrelated to the contexts in which they occur and in which raters may consider them (Problem complex §3 Mistaken dualistic views).	**Complexity theories (Axiom 1):** Conceive individuals as living organisms, especially emergence and inseparability of the study phenomena from their contexts (Problem complex §3 Mistaken dualistic views) as needed for investigations of causes and explanations (rather than just structural descriptions). Develop and use *contextualised approaches and methods* to study phenomena in their relevant contexts both within the organism and embedded into their larger external systems (e.g., abiotic, biotic, social, cultural), such as using tracking technology and methods of ambulatory [ecological momentary] monitoring (Problem complexes §3 Mistaken dualistic views and §4 Lack of definition and theoretical distinction of study phenomena). Apply *mereological concepts* to study whole–part relations (e.g., principle of emergence, catalysis, apperception). Develop theories of *processual change and development* and *dynamic causal relations* (e.g., using theories of complex dynamic systems and transdisciplinarity), not just analytic theories describing elemental structures isolated from the contexts in which they occur (Problem complex §7 Constructification).

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §6. Operationalism: Logical errors and impeded theory development		
A frequent idea in psychology is that a study phenomenon's meaning could be established through the operations needed for its investigation, manipulation or elicitation. This *logical error* conflates the study phenomena with the means of their investigation—psychologists' cardinal error. It entails further logical errors (e.g., *cause–effect conflation*, Table 1), toolbox thinking, proliferation of overlapping construct, and adapting research topics to the methods, rather than vice versa. Operationalism conflates disparate scientific activities (e.g., definition, hypothesis testing, data generation), thereby distorting scientific concepts and procedures. This impedes development of both theories about study phenomena and theories about methods.	Rating methods enable the flexible design of operational procedures, ad libitum and at low cost for any topic describable in small chunks of colloquial language ("*anything goes*" research attitude). This constitutes *a-theoretical instrumentalism* but enables the large-scale data generation needed to statistically analyse the outcomes of redundant operations for consistency (reliability) and empirical convergence that are considered to be meaningful for some reason or purpose (validity; Problem complex §11 Statisticism). Constructs are often defined through correlating item ratings (e.g., using factor analysis), thereby conflating scientific definition with empirical investigation. Belief rating 'scales' that are nominally associated with phenomena (e.g., 'anxiety scales') could be a valid method to study them (*toolbox thinking*).	**Anthropogenicity (Axiom 3):** Reflect on everyday constructs and lay definitions and their possible impact on the constructs and definitions developed and used in scientific psychology. Reflect on how researchers' practical engagement with and collective appraisal of reality influences the knowledge they develop about that reality (e.g., learn from the philosophers, historians and sociologists of science). Explore the ways in which a field's key concepts are theory-laden, socially embedded and historically contingent by explicating basic (hidden) assumptions and studying the concepts' history (i.e., using literature older than just 5-10 years). Develop philosophical and theoretical definitions specifying the essences of study phenomena (e.g., using philosophy of science approaches). Clearly distinguish these definitions from the operational procedures needed to empirically investigate the phenomena defined (procedurism). Advance both, theories about the study phenomena and theories about methods (Problem complexes §4 Lack of definition and theoretical distinction of study phenomena, §10 Quantificationism and §11 Statisticism).

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §7. Constructification: Studying constructs without also studying their intended referents		
Constructs are *conceptual systems* referring to entities (referents) considered to be meaningfully related but that actually (can) never occur all at once. Thus, constructs are only conceptual entities; this has three important implications: (1) Researchers can empirically study only some of a construct's referents chosen as *indicators*. (2) Construct referents can be all kinds of entities (e.g., abiotic, biotic, psychical, social, cultural). Despite differences in their forms of occurrence and accessibility thus requiring different research methods, heterogeneous referents can be conceptually integrated through abstraction into *blended constructs* (e.g., emphasising aspects considered relevant). (3) Given (1) and (2), constructs imply *more (surplus) meaning* than their concrete indicators, which therefore cannot be reflected in constructs as individuals can perceive them at a given moment. Psychologists' confusions arise from frequent *construct–referent conflation* (Table 1) and because constructs can be used as means of exploration (scientific constructs) but can also be explored in themselves (e.g., everyday constructs)—but, in a study, a given construct logically cannot be both. The constructs' construal on different levels of abstraction entails *nested conceptual structures* (symbolised with words) with complex meanings and referents. This may entail, especially in language-based research (Problem complex §8 Naïve use of language-based methods), that researchers study referents that are constructs in themselves rather than the actual concrete phenomena of interest to which these constructs refer and which therefore remain unstudied—the fallacy of *constructification*.	Inferential rating items describe not concrete phenomena but constructs that refer to (often heterogeneous kinds of) study phenomena in more general and abstract terms (Problem complex §4 Lack of definition and theoretical distinction of study phenomena). Their contents therefore can be judged also on the mere conceptual–semantic level (Problem complex §8 Naïve use of language-based methods). But even for judging (e.g., the intensity of) specific phenomena (e.g., specific behavioural acts), if considered, raters must make implicit comparisons over occasions, thus retrospective considerations, if not also over different individuals and different phenomena (e.g., other behaviours). Thus, ratings are inherently based on abstractions and generalisations. This entails the fallacy of constructification—researchers study only the everyday constructs that raters have developed *about* the actual phenomena of interest but not these phenomena (e.g., behavioural and psychical processes) in themselves, which therefore remain unstudied.	**Complexity theories (Axiom 1), Anthropogenicity (Axiom 3):** Metatheoretically and theoretically define the study phenomena (e.g., constructs, psyche, behaviours, language, actions), considering their processual, irreversible nature, momentariness, dynamicity, intra-individual variability, subjectivity, uniqueness, equifinality and multifinality and contextual embeddedness (Problem complexes §3 Mistaken dualistic views, §4 Lack of definition and theoretical distinction of study phenomena and §5 Reductionism). Develop methodologies, methods of data generation and methods of data analyses that are suitable for studying these phenomena and these particular properties (e.g., accessibility, forms and contexts of occurrence; Problem complexes §3 Mistaken dualistic views, §4 Lack of definition and theoretical distinction of study phenomena and §5 Reductionism). Develop concepts and theories of functional and causal relations between phenomena, thus of processual change and development (e.g., using complex dynamic systems and transdisciplinary theories), not just analytic and descriptive theories of elemental structures described on abstract levels (e.g., in everyday constructs), which preclude investigations of underlying processes (Problem complex §5 Reductionism).

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §8. Naïve use of language-based methods: Reification and studying merely linguistic propositions		
Language is often equated with verbal behaviour and psychical phenomena, overlooking its complexity (e.g., semantic networks, arbitrariness, variations, socio-cultural conventions), and its inherently representational and composite nature (signifier–referent–meaning relations). The role of semiotic systems for enabling conceptualisation and abstraction is hardly considered, therefore intricacies of language-based methods often overlooked. This entails fallacies, such that signifiers are equated with their meanings or their referents (psychologists' cardinal error; Table 1) or linguistic abstractions are reified as real concrete entities (e.g., 'traits'; fallacy of misplaced concreteness), which may mislead to take descriptions of the study phenomena for their explanation, resulting in explanatory circularity (Table 1). These fallacies also entail risks for studying merely linguistic propositions rather than the designated phenomena in themselves (Problem complex §7 Constructification).	Rating items are often thought to reflect standardised meanings (signifier–meaning conflation) or are equated with the phenomena they describe (signifier–referent conflation; Table 1). Variations in raters' (and researchers') item interpretation and use are often ignored although this entails that raters may consider in their ratings different meanings, thus also different phenomena than intended by researchers (Problem complex §1 Psychologists' own role in their research). Researchers often conceive item responses as verbal behaviours, mixing up raters' semantically guided meaning construction, everyday beliefs and hand movements for ticking off answer boxes (Problem complex §4 Lack of definition and theoretical distinction of study phenomena), leading to just pseudo-empirical findings. Rating-based research runs the risk of studying just linguistic propositions and the constructs designated (Problem complex §7 Constructification), both of which are often mistaken for the concrete phenomena to which they are intended to refer, thereby also often conflating description with explanation (Problem complexes §1 Psychologists' own role in their research and §4 Lack of definition and theoretical distinction of study phenomena).	**Complexity theories (Axiom 1), Anthropogenicity (Axiom 3):** Develop concepts and theories that conceptualise human psychical life as intrinsically mediated by signs and language as inseparable from its users' minds. This involves defining and differentiating language, psyche and behaviour to explore their interrelations, such as using complex dynamic system theories (Problem complex §4 Lack of definition and theoretical distinction of study phenomena) but also semiotic, semantic and pragmatic theories of language in exchange and collaboration with linguists. Develop and advance analytical approaches for language-based methods, which are inherently interpretive and cannot be standardised (e.g., using concepts of interpretive hermeneutics). Analyse textual materials produced in the researchee's own words and ideally in everyday contexts (e.g., open-ended surveys, diaries, social media communications) so that researchers' task is limited to selection and analysis of verbal materials (e.g., using semantic algorithms), without interfering in their production. Explore interpretations also on the individual level to avoid conducting mere semantic analyses and thus to study only codified conventional meaning structures (Problem complexes §11 Statisticism and §12 Nomotheticism).

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §9. Variable-based psychology and data-driven approaches: Overlooking the semiotic nature of 'data'		
'Data' are interpreted as either the phenomena to be studied or the variables and values that are to be analysed in lieu of the actual study phenomena. Conflating these two meanings entails logical errors in analyses (Problem complex §6 Operationalism) and interpretation (psychologists' cardinal error). The same conflation and logical errors occur for the term 'variables'. Without sufficient consideration of data as sign (semiotic) systems (Problem complex §8 Naïve use of language-based methods), researchers overlook that data are inherently theory-laden (e.g., reflected in beliefs about 'data-driven' approaches). This entails that methodologies for data generation are hardly developed. Lack of transparency in data generation (representation theorems) can entail mismatches with methods of data analysis (e.g., uniqueness theorems) and result interpretations (Problem complex §10 Quantificationism). Analyses of collective variables (encoding blended constructs with heterogeneous referents) mislead statisticians' result interpretations because it cannot be identified which referents and whether singly or collectively are relevant for a particular association found.	The dual function of rating 'scales' as description of the empirical study system and as symbolic (data) system promotes their conflation (psychologists' cardinal error) and leaves the specification of each system and their relations (representation theorems) to raters' implicit unknown decisions, thereby precluding transparency in data generation (and thus traceability; Problem complex §10 Quantificationism). Rating data are analysed only for the meanings that researchers consider (Problem complex §1 Psychologists' own role in their research), ignoring their semantic fields of meaning (Problem complex §8 Naïve use of language-based methods), raters' context-dependent meaning construction and particular interpretive perspective (Problem complex §3 Mistaken dualistic views), thus misinterpreting the information that particular rating data and the statistical results obtained from them can actually reflect.	**Complementarity (Axiom 2), Anthropogenicity (Axiom 3):** Develop a clear, unambiguous terminology, even if cumbersome, that avoids jingle–jangle and other fallacies and thus conflations. Develop methodologies of data generation (e.g., classes of methods determined by their modes of accessibility and forms of occurrence; Problem complexes §4 Lack of definition and theoretical distinction of study phenomena, §5 Reductionism and §7 Constructification) that are suited to establish transparency in the meaning of the data produced and their relations to their referents (e.g., data generation traceability, numerical traceability; Problem complex §10 Quantificationism). Carefully consider limitations in the interpretation of results from collective variables encoding (blended) constructs regarding their relations to their actual referents (Problem complex §7 Constructification).

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §10. Quantificationism: Numeralisation instead of measurement		
Quantification is widely believed to be essential for any science, a value in itself and generally better than qualitative findings—the belief of *quantificationism*. Still, elementary measurement concepts are largely unknown in psychology, such as *measurands*, *quantity* and *quality,* resulting in inadequate application of *quantity axioms.* Largely unknown are also concepts of *number* versus *numeral*, resulting in common *numeral–number conflation* (Table 1) and misguided interpretations of numerical data and their (quantitative) meanings. Operationalism entailed the erroneous belief that operations that yield convergent numerical results could provide evidence of quantitative properties in the study phenomena and could constitute valid 'instruments' for 'measuring' a construct (Problem complex §6 Operationalism). But traceable links from the measurand to both the result (*data generation traceability*) and a quantity of known defined meaning (*numerical traceability*) cannot be established, thus precluding the (1) results' justified attribution to the measurands, and their (2) publicly interpretable quantitative meaning regarding the property studied—the two *criteria of measurement* (implicitly) applied across the sciences. Measurement scales must fulfil *four methodological functions,* which cannot be substituted for each other: they serve as (1) 'instruments' enabling empirical interactions with study phenomena and properties; (2) structural data format; (3) conceptual data format; and (4) conventionally agreed reference quantities.	Raters are required to assign a broad range of quantitative information flexibly to a fixed, narrow range of values; thus, to adapt their judgements to the 'scale' rather than to the study phenomena by constructing *different* quantitative meanings for the *same* answer category, which can distort and even inverse quantitative relations. Still, researchers rigidly score the verbal answer values in always the same ways, replacing words with numerals. This recoding ignores the logico-semantic meanings that the verbal categories could actually have as well as raters' often rather trivial reasons for ticking off these categories, both of which are discordant with the quantitative meanings that researchers ascribe to the numerically recoded scores (Problem complex §1 Psychologists' own role in their research). This recoding also takes off all information about the properties ('x' *of what*?) and quantities (*how much* of that is 'x') to which the scores may refer. This misleads researchers to assume these scores could be interpreted flexibly as indicating quantities of any property as required for their research (e.g., agreement as a property of the study phenomena in themselves rather than as part of the judgement process; Problem complex §1 Psychologists' own role in their research). Steven's 'scale' categories lead psychologists to overlook that just implementing structural and conceptual data formats is not measurement. Without traceable relations to the measurands and known quantity references, rating 'scales' enable just *numeralisation*—the creation of numerical scores that are neither attributable to the measurands (e.g., individuals) nor publicly interpretable (i.e., *how much/many of what*).	**Complementarity (Axiom 2), Anthropogenicity (Axiom 3):** Develop both theories of measurement and theories of the study phenomena (Problem complexes §4 Lack of definition and theoretical distinction of study phenomena, §5 Reductionism, §7 Constructification and §9 Variable-based psychology and data-driven approaches) that incorporate the philosophy-of-science foundations of quality and quantity in order (1) to theoretically justify the possibility or impossibility of identifying quantitative (divisible) properties in psychical and behavioural phenomena (their *measurability*) and of specifying for them known quantities that are suitable as reference standards; and (2) to elaborate these quantitative properties' possible (ir)relevance to the given phenomena's functioning and development and their interrelations with other phenomena. Derive from these two types of theories quantitative methodologies, methods (e.g., using fuzzy categories as in behavioural coding) and models (e.g., based on computerised algorithms) that are adapted to the study phenomena and their particular properties and that allow traceable relations to the measurands and known quantity references to be established in empirical studies that can therefore yield results that are both attributable to measurands and publicly interpretable. Advance also non-quantitative methods of both data generation and data analysis to enable contextualised research on complex dynamic study phenomena (Problem complex §4 Lack of definition and theoretical distinction of study phenomena) as well as approaches of systematic and transparent interpretation (Problem complex §8 Naïve use of language-based methods), meta-synthesis and meta-theorising, which are needed for concept development. Profit from the expertise of the social, health and other sciences and their enormous portfolio of pertinent methodologies and methods many of which are still largely unknown in psychology.

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §11. Statisticism: Result-based data generation, methodomorphism and pragmatic quantification instead of measurement		
Without traceable relations to measurands and to known quantities, quantitative meaning for scores (Problem complex §10 Quantificationism) can be created only through between-case comparisons, making sample-level statistics essential for implementing quantitative approaches in psychology. But this means comparing scores with unknown quantity information to create quantitative meaning for them, which fails if all scores are the same and may be useful for pragmatic purposes only. Statistics is often regarded an end it itself and mistaken for the basis of science (and measurement)—a notion called *statisticism*. Statistics are based on theories and assumptions, which impose structures onto the data that cannot be separated from those of the study phenomena and that may, if erroneously attributed onto these latter, bias pertinent concepts and theories—the fallacy of *methodomorphism*. The more complex statistical methods are, the more obscured become the statistical results' relations to the actual study phenomena; this makes it difficult to scrutinise the adequacy of analytical tests and the appropriateness of interpretations. Sample-level statistics (e.g., effect sizes) are abstract parameters describing distribution patterns in a sample but they can neither be attributed to the samples' single measurands (e.g., single individuals' body height; Problem complex §12 Nomotheticism) nor create quantitative meaning for them (e.g., how tall is that?), thus cannot enable measurement.	To enable between-case comparisons, psychometricians develop rating 'scales' enabling the generation of scores that differentiate well (discrimination) and consistently (reliability) between cases and in ways considered meaningful (validity), such as by selecting items that produce norm-distributed values, show desired levels of item difficulty and item discrimination, or coherent score distributions across different items used for the same construct. But this adapts methods and results to statistical criteria and theories rather than to properties of the actual study phenomena (Problem complex §1d Psychologists' own role in their research); thus enabling only *result-dependent data generation* but not measurement. Psychological validity concepts concern relations between phenomena of *different* qualities (e.g., those described in items used for different, theoretically (un-)related constructs or real-world outcomes like health or job performances; Problem complex §6 Operationalism). By contrast, measurement is about capturing *quantitative (divisible)* properties *of one specific defined quality.*	**Complexity theories (Axiom 1), Anthropogenicity (Axiom 3):** Use the methodological rationales underlying the principles of data generation traceability and numerical traceability, specifying the relations between empirical and symbolic study system (Problem complex §10 Quantificationism), to develop analogous principles of *data analysis traceability* specifying general rationales underlying the transformations that different kinds of analytical methods make within the symbolic study system (e.g., abstracting from uniqueness theorems and specific statistical theories). Such general principles will help establish transparency in the analytical results' relations to the original raw data with regard to the information that these reflect about the measurands and their qualitative and quantitative meanings (e.g., rationales for grouping cases and choosing units of aggregation), thereby guiding researchers' result interpretation with a clear focus on the empirical study system (Problem complex §12 Nomotheticism; Figures 6, 8). Use and further develop simpler statistical procedures that remove themselves only slightly from the original data, enabling meaningful interpretation regarding the samples (e.g., groups of individuals or repeated observations of single individuals) analysed (Problem complex §12 Nomotheticism). Linear analyses of sample-level convergence (e.g., factorial analysis), by contrast, depart very far from the original data and involve more assumptions that are not explicitly considered in the formal model and result interpretation and that cannot explore the nonlinear relations found in living organisms. Apply and advance knowledge of qualitative mathematics and other models needed to analyse dynamic processes.

(Continued)

10.3389/fpsyg.2022.1009893

TABLE 2 (Continued)

Metatheoretical and methodological problem complexes	Implementation in rating 'scales'	Axioms, theoretical concepts, methodologies and methods required to tackle these problems
Problem complex §12. Nomotheticism: Sociological/ergodic fallacy and primacy of sample-based over case–by–case based nomothetic approaches		
Many psychologists erroneously believe that structures of *inter*-individual differences could be informative about *intra*-individual functioning and development. Therefore, *sample-based nomothetic* (variable-oriented) approaches are widely used in which individuals are grouped by properties that the researchers find a-priori meaningful (Problem complexes §1 Psychologists' own role in their research and §11 Statisticism) and the aggregates of the groups thus-created are generalised to single individuals. But such inferences build on the *sociological/ergodic fallacy* because, in phenomena that change and develop, diachronic and synchronic variations are not isomorphic. Valid sample–to–individual inferences would logically require the assumptions of psychical homogeneity and stationarity—but these contradict the empirical data bases as well as fundamental design principles of complex living systems: many–to–one (degeneracy, equifinality) and one–to–many structure–function relations (pluripotency, multifinality). In *case–by–case based nomothetic* (individual-oriented) approaches, by contrast, individuals are grouped by their commonalities and differences in the study phenomena—by what is indeed common to all cases. Degeneracy and pluripotency can be studied by exploring these groups further for communalities and differences in their underlying intra-individual structures and processes, thereby linking individuals with theory development.	The differential focus is inherent to rating 'scales' because it is needed to create quantitative meaning for rating scores (Problem complexes §10 Quantificationism and §11 Statisticism). Therefore, sample-based nomothetic approaches have become the default approach for analysing rating data, which contributed to their primacy in psychological research. Sample-level aggregates and their structures are commonly attributed to the single individuals and erroneously used to derive theories about individual-level phenomena (e.g., in Five Factor theory in 'personality' research, between-individual differences are conceptualised as an explanation of intra-individual phenomena), ignoring that this is based on the ergodic fallacy and conflating description with explanation (Problem complex §8 Naïve use of language-based methods; Table 1).	Advance and develop methods that are suited to explore *intra-individual* processes, change and development (Problem complex §11 Statisticism), both methods of data collection (e.g., momentary and situated recording of behaviours, physiological responses and experiential reports in ambulatory monitoring) as well as methods of data analysis (e.g., individual-oriented approaches like configurational frequency analysis; processual analyses) that allow researchers to adequately consider the non-ergodicity of psychical and behavioural phenomena. These should be integrated into suitable individual–socio–ecological frames of reference that need to be developed for the contextualised in-depth exploration of individuals using (instead of inductive differential generalisation from large samples) abductive generalisation to create meaningful findings that allow researchers to develop theories about individuals, their functioning and development.

For a better overview, relevant references are not included here but provided in the main text.

10.3389/fpsyg.2022.1009893

contexts in general (Fleck, 1935; Kuhn, 1962) and in one's own field of research in particular. When some psychology journals demanded citations to be limited to publications from just the last 5–10 years, the history of thought of key psychological concepts, theories and methods got partly out of sight. Knowing more about their origins and the contexts in which they had had once been created, will empower psychologists to critically reflect on the (implicit) foundations of their established research practices.

Make explicit and elaborate the own metatheory and an unambiguous terminology

To establish psychology as a science, psychologists must take intellectual responsibility for its metatheoretical foundations (Vautier et al., 2014). Critical reflection and controversial debate presuppose that basic assumptions are made explicit (Axiom 3). This is a laborious task. It requires precise definitions, terminologies and concepts that cannot build on everyday language and that enable the crucial distinction between the study phenomena and the means of their exploration—thus, to avoid psychologists' cardinal error. Such conceptual developments take far more efforts and time than empirical studies, on which currently most psychologists focus their research activities. But this conceptual research is necessary to identify inconsistencies and mismatches (see, e.g., Lilienfeld et al., 2015; Slaney and Garcia, 2015; Uher, 2021c,d), to trace theories and concepts to their origins, and to scrutinise their (meanwhile) often hidden underlying assumptions in order to enable reappraisal, critical reflection and even radical change and renewal (Danziger, 1990; Bennett and Hacker, 2003; Valsiner, 2012; Fahrenberg, 2013, 2015).

Develop theories and concepts of the study phenomena

With rating 'scales', psychologists implemented a simplified but appealing image of natural science from which they created an equally simplified and appealing image of psychology as a science, but which cannot meet the complexities of its subject matter (Mazur and Watzlawik, 2016). We need theories that allow us to conceive individuals as living beings, as open self-organising systems featuring complementary phenomena (Axiom 2) and dynamic interrelations across their multi-layered systemic contexts (Axiom 1)—that is, theories not simply of elemental properties and structures but of processes, relations, dynamicity, subjectivity, emergence, catalysis and transformation (Fahrenberg, 2013, 2015; Cabell and Valsiner, 2014; Salvatore, 2015; Toomela, 2021; Valsiner, 2021). To explore continuous, dynamic, unprecedented and creative change and development, we need not simplistic dualistic but inclusive concepts (Table 2; Valsiner, 1997) as well as dialogic and dialectic theories (Veraksa et al., 2022).

Develop methodology—The philosophy and theories of research methods

Psychology has become, in parts, a mere craft—focussed on the technicalities of data analyses (Maslow, 1946; Brower, 1949; Toomela and Valsiner, 2010). It is replete with theories of statistical analysis but devoid of theories of quantitative data generation (Uher, 2018a, 2021a). A quantitative methodology, specifying rationales and basic principles for linking the study phenomena with formal and quantitative models of investigation is still missing—the principles of data generation traceability and numerical traceability (Uher, 2020b, 2022) can be just a start (e.g., principles of data *analysis* traceability; Table 2).

Psychologists have still hardly considered the philosophical underpinnings of what quality and quantity actually are and how they are related (Hartmann, 1964; Rudolph, 2013; Uher, 2018a). Their philosophy-of-science definition highlights that quantification is useful for exploring only a minority of psychology's study phenomena. Change and development, the key characteristics of most of the phenomena studied in psychology, are not just more of the same but involve *qualitatively different* structures and processes (Axioms 1)—therefore, quantification is of limited value (Valsiner, 2017b). Rather than sample-based quantitative analysis and meta-analysis for exploring group-averages, we need case–by–case based nomothetic approaches to explore individual functioning and development (Salvatore and Valsiner, 2010) as well as methods of qualitative synthesis and meta-synthesis, which are also essential for concept and theory development (Sim and Mengshoel, 2022).

Language is essential for science because results can be interpreted and communicated only in language (Axiom 3; Wittgenstein, 1922; Bohr, 1937). Language is also central for studying many psychical phenomena (Vygotsky, 1962; Valsiner, 2001; Salvatore et al., 2019). Language-based methods will always be important means of psychological investigation. Therefore, psychologists should acquire at least some basic knowledge of semiotics and semantics. Indeed, semantic computer algorithms, for example, are efficient methods to analyse open-ended verbal responses that can replace rating 'scales' in large-scale inquiries (Arnulf et al., 2021; Smedslund, 2021).

Replicability of psychological interventions and their effectiveness is important. But current approaches to validity and replicability are just pragmatic, providing evidence only of utility. In lack of traceability to the study phenomena, they provide no explanations or theories why these interventions are useful and how their effects come about (Uher, 2021c,d). Without understanding the phenomena, the actual causes of repeatable findings—which may be completely unrelated to the study phenomena—cannot be explored. *Explanatory psychology* exploring the microgenetic, ontogenetic and phylogenetic development of its study phenomena requires identification of abstract principles (Valsiner, 2021), not just endless repetitions of scores produced in ways that remain opaque and lay psychological. Instead of merely accumulating empirical findings, crucial

experiments the results of which require major theoretical changes are particularly illuminating (Valsiner, 2014).

Develop new research methods appropriately adapted to the study phenomena

Scientific discoveries of lasting and stimulating value have always led to new method developments. Psychologists cannot hope to progress by continuously re-applying old techniques to new, unprecedented problems (Brower, 1949). Methods of data generation must be developed that are appropriate to the study phenomena's particular modes of accessibility to researchers and researchees and that consider possible complementary relations (Axiom 2; Table 2; Uher, 2019). Suitable methods are needed both to capture and to analyse psychical phenomena's key features, in particular, their intra-individual variation, dynamicity, ephemerality, irreversibility, uniqueness, subjectivity, equifinality and multifinality (Sato et al., 2009; Valsiner, 2017b).

Psychologists' present knowledge base of mathematics is limited and outdated. They need to widen their considerations to mathematical systems that are suitable to arrive at generalised knowledge about complex dynamic phenomena (e.g., quantum probability, topology, knot theory; Rudolph, 2013). To enable measurement, processes must be devised that establish traceable relations both to the phenomena and properties under study and to their quantitative and qualitative meanings. This involves also elaborating—and respecting—inherent limitations in the measurability of many psychological study phenomena (Uher, 2020b, 2022).

Learn from other disciplines and their advancements

The complex phenomena of psyche, behaviour and society play central roles in all domains of individuals' lives. Psychology's focus on individuals and these phenomena puts the discipline at the intersection of many other sciences and of philosophy (Uher, 2021b). Therefore, psychologists are uniquely positioned to collaborate with other disciplines and to learn from the philosophical perspectives, metatheories, methodologies and methods that they have developed from their particular perspectives and for their particular research questions. Transdisciplinary research plays an important role in these endeavours because it aims to develop unitary frameworks that transcend disciplinary boundaries (e.g., complexity theories, dialectic theories). It can therefore highlight connections, differences and communalities across sciences—and thus, promising starting points for cross-scientific exchange and collaboration.

> "The list of mistakes presented here was not intended to be exhaustive, nor the proposed solutions to encompass all possibilities. The present aim was, merely, to alert colleagues about the existence of these fallacies, and to provide them with a source and a reference. Hopefully this work will help prevent perpetuating these … mistakes on the grounds that 'this is how things have always been done' and 'no-one ever said it was wrong.' Now, you know" (Vanrullen, 2011, p. 6).

Data availability statement

The original contributions presented in this research are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This research was funded by a Marie Curie Fellowship of the European Commission's FP7 Programme awarded to me (EC grant agreement number 629430).

Acknowledgments

The author thanks the two reviewers for their constructive feedback.

Conflict of Interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Abran, A., Desharnais, J.-M., and Cuadrado-Gallego, J. J. (2012). Measurement and quantification are not the same: ISO 15939 and ISO 9126. *J. Software Evol. Process* 24, 585–601. doi: 10.1002/smr.496

AERA, APA, and NCME (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

Allport, G. W., and Odbert, H. S. (1936). Trait names: a psycholexical study. *Psychol. Monogr.* 47, 1–171.

Althusser, L., and Balibar, E. (1970). *Reading Capital*. London: New Left Books.

Anderson, C. A., Allen, J. J., Plante, C., Quigley-McBride, A., Lovett, A., and Rokkum, J. N. (2018). The MTurkification of social and personality psychology. *Personal. Soc. Psychol. Bull.* 45, 842–850. doi: 10.1177/0146167218798821

Arnulf, J. K. (2020). "Wittgenstein's revenge: how semantic algorithms can help survey research escape Smedslund's labyrinth" in *Respect for Thought*. eds. T. G. Lindstad, E. Stänicke and J. Valsiner (Cham: Springer International Publishing), 285–307.

Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., and Bong, C. H. (2014). Predicting survey responses: how and why semantics shape survey statistics on organizational behaviour. *PLoS One* 9:e106361. doi: 10.1371/journal.pone.0106361

Arnulf, J. K., Larsen, K. R., Martinsen, Ø. L., and Nimon, K. F. (2021). Semantic algorithms in the assessment of attitudes and personality. *Front. Psychol.* 12:720559. doi: 10.3389/fpsyg.2021.720559

Asch, S. E. (1955). Opinions and social pressure. *Sci. Am.* 193, 31–35. doi: 10.1038/scientificamerican1155-31

Ashworth, P. D. (2009). William James's "psychologist's fallacy" and contemporary human science research. *Int. J. Qual. Stud. Health Well Being* 4, 195–206. doi: 10.3109/17482620903223036

Bandura, A. (1996). Ontological and epistemological terrains revisited. *J. Behav. Ther. Exp. Psychiatry* 27, 323–345. doi: 10.1016/S0005-7916(96)00049-3

Bandura, A. (2006). Toward a psychology of human agency. *Perspect. Psychol. Sci.* 1, 164–180. doi: 10.1111/j.1745-6916.2006.00011.x

Barrett, P. (2003). Beyond psychometrics. *J. Manag. Psychol.* 18, 421–439. doi: 10.1108/02683940310484026

Barrett, P. (2018). The EFPA test-review model: when good intentions meet a methodological thought disorder. *Behav. Sci.* 8:5. doi: 10.3390/bs8010005

Bartlett, F. C. (1932/1995). *Remembering: A study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Baumeister, R. F., Vohs, K. D., and Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: whatever happened to actual behavior? *Perspect. Psychol. Sci.* 2, 396–403. doi: 10.1111/j.1745-6916.2007.00051.x

Benjamin, A. C. (1955). *Operationism*. Springfield, IL: Charles C. Thomas.

Bennett, M. R., and Hacker, P. M. S. (2003). *Philosophical Foundations of Neuroscience*. Oxford: Wiley-Blackwell.

Bergman, L. R., and Trost, K. (2006). The person-oriented versus the variable-oriented approach: are they complementary, opposites, or exploring different worlds? *Merrill-Palmer Q.* 52, 601–632. doi: 10.1353/mpq.2006.0023

Bickhard, M. H. (2001). The tragedy of operationalism. *Theory Psychol.* 11, 35–44. doi: 10.1177/0959354301111002

Biesanz, J. C., and Human, L. J. (2010). The cost of forming more accurate impressions: accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers qithout an explicit goal. *Psychol. Sci.* 21, 589–594. doi: 10.1177/0956797610364121

BIPM (2006). *BIPM: The International System of Units (SI)*. (8th Edn.). Available at: http://www.bipm.org/

Bohr, N. (1937). Causality and complementarity. *Philos. Sci.* 4, 289–298. doi: 10.1086/286465

Borkenau, P., and Müller, B. (1991). Breadth, bandwidth, and fidelity of personality-descriptive categories. *Eur. J. Personal.* 5, 309–322. doi: 10.1002/per.2410050404

Borsboom, D. (2005). *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*. Cambridge: Cambridge University Press.

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., and Franić, S. (2009). "The end of construct validity" in *The Concept of Validity: Revisions, New Directions, and Applications* (Charlotte, NC: IAP Information Age Publishing), 135–170.

Borsboom, D., and Mellenbergh, G. J. (2004). Why psychometrics is not pathological. *Theory Psychol.* 14, 105–120. doi: 10.1177/0959354304040200

Borsboom, D., and Scholten, A. Z. (2008). The Rasch model and conjoint measurement theory from the perspective of psychometrics. *Theory Psychol.* 18, 111–117. doi: 10.1177/0959354307086925

Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: Macmillan.

Brigandt, I., and Love, A. (2022). "Reductionism in biology," in *The Stanford Encyclopedia of Philosophy*. ed. E. N. Zalta. Available at: https://plato.stanford.edu/archives/sum2022/entries/reduction-biology/

Brower, D. (1949). The problem of quantification of psychological science. *Psychol. Rev.* 56, 325–331. doi: 10.1037/h0061802

Buss, D. M. (2009). How can evolutionary psychology successfully explain personality and individual differences? *Perspect. Psychol. Sci.* 4, 359–366. doi: 10.1111/j.1745-6924.2009.01138.x

Cabell, K. R., and Valsiner, J. (2014). *The Catalyzing Mind. Beyond Models of Causality*. New York: Springer.

Campbell, D. T. (1974). "Qualitative knowing in action research. Kurt Lewin award address," in *Society for the Psychological Study of Social Issues, Presented at the Meeting of the American Psychological Association*. New Orleans, LA.

Campbell, D. T., and Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56, 81–105. doi: 10.1037/h0046016

Capra, F. (1997). *The Web of Life: A New Synthesis of Mind and Matter*. New York: Anchor Books.

Chen, M., Regenwetter, M., and Davis-Stober, C. P. (2020). Collective choice may tell nothing about anyone's individual preferences. *Decis. Anal.* 18, 1–24. doi: 10.1287/deca.2020.0417

Collingwood, R. G. (1940). *An Essay on Metaphysics*. UK Oxford University Press.

Condon, D., Beck, E., and Jackson, J. (2021). Age differences in personality structure. *Innov. Aging* 5:564. doi: 10.1093/geroni/igab046.2153

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *Am. Psychol.* 12, 671–684. doi: 10.1037/h0043943

Cronbach, L. J., and Meehl, P. E. (1955). Construct validity in psychological tests. *Psychol. Bull.* 52, 281–302. doi: 10.1037/h0040957

Czichos, H. (2011). "Introduction to metrology and testing" in *Springer Handbook of Metrology and Testing* (Berlin, Heidelberg: Springer), 3–22.

Danziger, K. (1979). The positivist repudiation of Wundt. *J. Hist. Behav. Sci.* 15, 205–230. doi: 10.1002/1520-6696(197907)15:3<205::AID-JHBS2300150303>3.0.CO;2-P

Danziger, K. (1985a). The methodological imperative in psychology. *Philos. Soc. Sci.* 15, 1–13. doi: 10.1177/004839318501500101

Danziger, K. (1985b). The origins of the psychological experiment as a social institution. *Am. Psychol.* 40, 133–140. doi: 10.1037/0003-066X.40.2.133

Danziger, K. (1990). *Constructing the Subject: Historical Origins of Psychological Research*. New York, NY: Cambridge University Press.

Danziger, K. (1997). *Naming the Mind: How Psychology Found its Language*. London, UK: Sage.

Danziger, K., and Dzinas, K. (1997). How psychology got its variables. *Can. Psychol./Psychologie Canadienne* 38, 43–48. doi: 10.1037/0708-5591.38.1.43

Dawes, R., Faust, D., and Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science* 243, 1668–1674. doi: 10.1126/science.2648573

De Ruiter, N. M. P., Van Der Steen, S., Den Hartigh, R. J. R., and Van Geert, P. L. C. (2017). Capturing moment-to-moment changes in multivariate human experience. *Int. J. Behav. Dev.* 41, 611–620. doi: 10.1177/0165025416651736

Dennett, D. (2012). A perfect and beautiful machine': what Darwin's theory of evolution reveals about artificial intelligence. *The Atlantic, June 22*. Available at: http://hdl.handle.net/10427/000489

Deutscher, G. (2006). *The Unfolding of Language: The Evolution of Mankind's Greatest Invention*. London, UK: Arrow.

Dewey, J. (1946). Peirce's theory of linguistic signs, thought, and meaning. *J. Philos.* 43, 85–95. doi: 10.2307/2019493

Diez Roux, A. V. (2002). A glossary for multilevel analysis. *J. Epidemiol. Community Health* 56:588 LP – 594. doi: 10.1136/jech.56.8.588

Duncan, O. D. (1984). *Notes on Social Measurement: Historical and Critical*. Nova York: Russell Sage Foundation.

Fahrenberg, J. (1979). The complementarity principle in psychophysiological research and somatic medicine. *Z. Klin. Psychol. Psychother.* 27, 151–167. PMID: 539037

Fahrenberg, J. (2002). *Psychologische interpretation. Biographien – Texte – Tests*. Bern: Hans Huber. ebook.

Fahrenberg, J. (2003). "Interpretation in psychology and social science—new approach or a neglected tradition?," in *Forum qualitative Sozialforschung/forum: qualitative social research, 4(2 SE-FQS Debate: Quality of Qualitative Research, Art. 45)*, 1–12.

Fahrenberg, J. (2013). *Zur Kategorienlehre der Psychologie: Komplementaritätsprinzip; Perspektiven und Perspektiven-Wechsel [on the category theory of psychology: Principle of complementarity, perspectives and changes of perspectives]*. Lengerich, Germany: Pabst Science Publishers.

Fahrenberg, J. (2015). *Theoretische Psychologie – Eine Systematik der Kontroversen*. Lengerich, Germany: Pabst Science Publishers.

Fahrenberg, J. (2019). *Wilhelm Wundt (1832–1920). Introduction, Quotations, Reception, Commentaries, Attempts at Reconstruction*. Lengerich: Pabst Science Publishers.

Fahrenberg, J., Myrtek, M., Pawlik, K., and Perrez, M. (2007). Ambulatory assessment – monitoring behavior in daily life settings. *Eur. J. Psychol. Assess.* 23, 206–213. doi: 10.1027/1015-5759.23.4.206

Faust, D. (2012). *Ziskin's Coping with Psychiatric and Psychological Testimony*. New York, US: Oxford University Press.

Feest, U. (2005). Operationism in psychology: what the debate is about, what the debate should be about. *J. Hist. Behav. Sci.* 41, 131–149. doi: 10.1002/jhbs.20079

Finkelstein, L. (2003). Widely, strongly and weakly defined measurement. *Measurement* 34, 39–48. doi: 10.1016/S0263-2241(03)00018-6

Fisher, W. P. (2009). Invariance and traceability for measures of human, social, and natural capital: theory and application. *Measurement* 42, 1278–1287. doi: 10.1016/j.measurement.2009.03.014

Flake, J. K., and Fried, E. I. (2020). Measurement schmeasurement: questionable seasurement practices and how to avoid them. *Adv. Methods Pract. Psychol. Sci.* 3, 456–465. doi: 10.1177/2515245920952393

Fleck, L. (1935). "Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv" in *Frankfurt am Main: Suhrkamp. English translation: (1979). The Genesis and Development of a Scientific Fact*. eds. T. J. Trenn and R. K. Merton (Chicago: University of Chicago Press)

Flynn, J. R. (2012). *Are We Getting Smarter? Rising IQ in the Twenty-First Century*. Cambridge: Cambridge University Press.

Franz, D. J. (2021). "Are psychological attributes quantitative?" is not an empirical question: conceptual confusions in the measurement debate. *Theory Psychol.* 32, 131–150. doi: 10.1177/09593543211045340

Funder, D. C. (1995). On the accuracy of personality judgment: a realistic approach. *Psychol. Rev.* 102, 652–670. doi: 10.1037/0033-295X.102.4.652

Gergen, K. J. (1985). The social constructionist movement in modern psychology. *Am. Psychol.* 40, 266–275. doi: 10.1037/0003-066X.40.3.266

Gergen, K. J. (2001). Psychological science in a postmodern context. *Am. Psychol.* 56, 803–813. doi: 10.1037/0003-066X.56.10.803

Giordani, A., and Mari, L. (2014). "Modeling measurement: error and uncertainty" in *Error and Uncertainty in Scientific Practice*. eds. M. Boumans, G. Hon and A. Peterson (Cambridge, UK: Pickering & Chatto), 79–96.

Green, C. D. (2001). Operationism again: what did Bridgman say? What did Bridgman need? *Theory Psychol.* 11, 45–51. doi: 10.1177/0959354301111003

Haggerty, K., and Porter, T. (1997). Trust in numbers: the pursuit of pbjectivity in science and public life. *Can. J. Sociol./Cahiers Canadiens de Sociologie* 22:279.

Haig, B. D., and Borsboom, D. (2008). On the conceptual foundations of psychological measurement. *Meas. Interdiscip. Res. Perspect.* 6, 1–6. doi: 10.1080/15366360802035471

Hanel, P. H. P., and Vione, K. C. (2016). Do student samples provide an accurate estimate of the general public? *PLoS One* 11:e0168354. doi: 10.1371/journal.pone.0168354

Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., and Ioannidis, J. P. A. (2022). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspect. Psychol. Sci.* 17, 239–251. doi: 10.1177/1745691620979806

Hartmann, N. (1964). *Der Aufbau der realen welt. Grundriss der allgemeinen Kategorienlehre [the Structure of the Real World. Outline of the General Theory of Categories]* (*3rd*). Berlin: Walter de Gruyter.

Hibberd, F. J. (2019). What is scientific definition? *J. Mind Behav.* 40, 29–52.

Hölder, O. (1901). Die Axiome der Quantität und die Lehre vom Mass (Band 53). Leipzig: Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch—Physische Classe.

Hyde, J. S. (2005). The gender similarities hypothesis. *Am. Psychol.* 60, 581–592. doi: 10.1037/0003-066X.60.6.581

James, W. (1890). *Principles of Psychology* (Vol. *1*). New York, NY: Holt.

JCGM100:2008 (2008). *Evaluation of measurement data – Guide to the expression of uncertainty in measurement (GUM)*. Joint Committee for Guides in Metrology (originally published in 1993). Available at: http://www.bipm.org/en/publications/guides/gum.html

JCGM200:2012 (2012). *International vocabulary of metrology – basic and general concepts and associated terms, 3rd Edn*. Paris: Joint Committee for Guides in Metrology.

John, O. P., Angleitner, A., and Ostendorf, F. (1988). The lexical approach to personality: a historical review of trait taxonomic research. *Eur. J. Personal.* 2, 171–203. doi: 10.1002/per.2410020302

John, O. P., and Robins, R. W. (2022). *Handbook of personality: Theory and research* (4th ed.). New York: Guilford.

Johnson, H. M. (1943). Index-numerology and measures of impairment. *Am. J. Psychol.* 56, 551–558. doi: 10.2307/1417354

Kane, M. T. (2013). Validating the interpretations and uses of test scores. *J. Educ. Meas.* 50, 1–73. doi: 10.1111/jedm.12000

Kant, I. (1781/1998) in *Kritik der reinen Vernunft*. ed. J. Timmermann (Hamburg: Felix Meiner Verlag)

Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. Scranton, PA: Chandler Publishing Co.

Kelley, T. L. (1927). *Interpretation of Educational Measurements*. Yonkers, NY: World.

Kelly, G. (1955). *The Psychology of Personal Constructs (volume 1 & 2)*. London, UK: Routledge.

Kenny, D. A. (1991). A general model of consensus and accuracy in interpersonal perception. *Psychol. Rev.* 98, 155–163. doi: 10.1037/0033-295X.98.2.155

Khanam, S. A., Liu, F., and Chen, Y.-P. P. (2019). Comprehensive structured knowledge base system construction with natural language presentation. *HCIS* 9:23. doi: 10.1186/s13673-019-0184-7

Klages, L. (1926). Grundlagen der Charakterkunde [The science of character; W.H. Johnston, Trans. 1932). Available at: http://dispater.atspace.com/

Kothari, C. R. (2004). *Research Methodology: Methods and Techniques* (*2nd ed.*). New Delhi: New Age International Publishers.

Krantz, D., Luce, R. D., Tversky, A., and Suppes, P. (1971). *Foundations of Measurement Volume I: Additive and Polynomial Representations*. San Diego: Academic Press.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kyngdon, A. (2011). Plausible measurement analogies to some psychometric models of test performance. *Br. J. Math. Stat. Psychol.* 64, 478–497. doi: 10.1348/2044-8317.002004

Laajaj, R., Macours, K., Pinzon Hernandez, D. A., Arias, O., Gosling, S. D., Potter, J., et al. (2019). Challenges to capture the big five personality traits in non-WEIRD populations. Science. *Advances* 5:eaaw5226. doi: 10.1126/sciadv.aaw5226

Lahlou, S. (1996). The propagation of social representations. *J. Theory Soc. Behav.* 26, 157–175. doi: 10.1111/j.1468-5914.1996.tb00527.x

Lahlou, S. (1998). *Penser Manger*. Paris, France: Les Presses Universitaires de France.

Lamiell, J. (1998). "Nomothetic" and "idiographic": contrasting Windelband's understanding with contemporary usage. *Theory Psychol.* 8, 23–38. doi: 10.1177/0959354398081002

Lamiell, J. (2003). *Beyond Individual and Group Differences: Human Individuality, Scientific Psychology, and William Stern's Critical Personalism*. Thousand Oaks, CA: Sage Publications.

Lamiell, J. (2013). Statisticism in personality psychologists' use of trait constructs: what is it? How was it contracted? Is there a cure? *New Ideas Psychol.* 31, 65–71. doi: 10.1016/j.newideapsych.2011.02.009

Lamiell, J. (2018). From psychology to psychodemography: how the adoption of population-level statistical methods transformed psychological science. *Am. J. Psychol.* 131, 471–475. doi: 10.5406/amerjpsyc.131.4.0471

Lamiell, J. (2019). *Psychology's Misuse of Statistics and Persistent Dismissal of its Critics*. Palgrave Macmillan Cham.

LaPiere, R. T. (1934). Attitudes vs. actions. *Soc. Forces* 13, 230–237. doi: 10.2307/2570339

Law, J. (2004). *After Method: Mess in Social Science Research*. New York, NY: Routledge.

Lewin, K. (1936). *Principles of Topological Psychology*. New York, NY: McGraw-Hill.

Likert, R. (1932). A technique for the measurement of attitudes. *Arch. Psychol.* 22, 1–55.

Lilienfeld, S. O., Sauvigné, K. C., Lynn, S. J., Cautin, R. L., Latzman, R. D., and Waldman, I. D. (2015). Fifty psychological and psychiatric terms to avoid: a list of inaccurate, misleading, misused, ambiguous, and logically confused words and phrases. *Front. Psychol.* 6:1100. doi: 10.3389/fpsyg.2015.01100

Locke, J. (1999). *An Essay Concerning Human Understanding*. Pennsylvania: The Pennsylvania State University.

Lovasz, N., and Slaney, K. L. (2013). What makes a hypothetical construct "hypothetical"? Tracing the origins and uses of the 'hypothetical construct' concept in psychological science. *New Ideas Psychol.* 31, 22–31. doi: 10.1016/j.newideapsych.2011.02.005

Ludeke, S. G., and Larsen, E. G. (2017). Problems with the big five assessment in the world values survey. *Personal. Individ. Differ.* 112, 103–105. doi: 10.1016/j.paid.2017.02.042

Lundmann, L., and Villadsen, J. W. (2016). Qualitative variations in personality inventories: subjective understandings of items in a personality inventory. *Qual. Res. Psychol.* 13, 166–187. doi: 10.1080/14780887.2015.1134737

Malich, L., and Rehmann-Sutter, C. (2022). Metascience is not enough – a plea for psychological humanities in the wake of the replication crisis. *Rev. Gen. Psychol.* 26, 261–273. doi: 10.1177/10892680221083876

Maraun, M. D., and Gabriel, S. M. (2013). Illegitimate concept equating in the partial fusion of construct validation theory and latent variable modeling. *New Ideas Psychol.* 31, 32–42. doi: 10.1016/j.newideapsych.2011.02.006

Maraun, M. D., and Halpin, P. F. (2008). Manifest and latent variates. *Meas. Interdiscip. Res. Perspect.* 6, 113–117. doi: 10.1080/15366360802035596

Maraun, M. D., Slaney, K. L., and Gabriel, S. M. (2009). The Augustinian methodological family of psychology. *New Ideas Psychol.* 27, 148–162. doi: 10.1016/j.newideapsych.2008.04.011

Mari, L. (2013). A quest for the definition of measurement. *Measurement* 46, 2889–2895. doi: 10.1016/j.measurement.2013.04.039

Mari, L., Carbone, P., Giordani, A., and Petri, D. (2017). A structural interpretation of measurement and some related epistemological issues. *Stud. Hist. Phil. Sci.* 65–66, 46–56. doi: 10.1016/j.shpsa.2017.08.001

Mari, L., Carbone, P., and Petri, D. (2015). "Fundamentals of hard and soft measurement" in *Modern Measurements: Fundamentals and Applications*. eds. A. Ferrero, D. Petri, P. Carbone and M. Catelani (Hoboken, NJ: John Wiley & Sons), 203–262. doi: 10.1002/9781119021315.ch7

Marsico, G., Andrisano Ruggieri, R., and Salvatore, S. (2015). *Reflexivity and Psychology*. Charlotte, NC: Information Age Publishing.

Maslow, A. H. (1946). Problem-centering vs. means-centering in science. *Philos. Sci.* 13, 326–331. doi: 10.1086/286907

Mason, P. H. (2010). Degeneracy at multiple levels of complexity. *Biol. Theory* 5, 277–288. doi: 10.1162/BIOT_a_00041

Mason, P. H. (2015). Degeneracy: demystifying and destigmatizing a core concept in systems biology. *Complexity* 20, 12–21. doi: 10.1002/cplx.21534

Maul, A. (2017). Rethinking traditional methods of survey validation. *Meas. Interdiscip. Res. Perspect.* 15, 51–69. doi: 10.1080/15366367.2017.1348108

Mazur, L. B., and Watzlawik, M. (2016). Debates about the scientific status of psychology: looking at the bright side. *Integr. Psychol. Behav. Sci.* 50, 555–567. doi: 10.1007/s12124-016-9352-8

McCrae, R. R. (2009). The physics and chemistry of personality. *Theory Psychol.* 19, 670–687. doi: 10.1177/0959354309341928

McCrae, R. R., and Costa, P. T. J. (1997). Personality trait structure as a human universal. *Am. Psychol.* 52, 509–516. doi: 10.1037/0003-066X.52.5.509

McCrae, R. R., and Costa, P. T. Jr. (2008). "The five-factor theory of personality" in *Handbook of Personality: Theory and Research*. eds. O. P. John, R. W. Robins and L. A. Pervin. *3rd* ed (New York, NY: Guilford Press), 159–181.

McKennell, A. C. (1974). Surveying attitude structures: a discussion of principles and procedures. *Qual. Quant.* 7, 203–294. doi: 10.1007/BF00153932

Mehl, M. R. (2017). The electronically activated recorder (EAR): a method for the naturalistic observation of daily social behavior. *Curr. Dir. Psychol. Sci.* 26, 184–190. doi: 10.1177/0963721416680611

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *Am. Psychol.* 50, 741–749. doi: 10.1037/0003-066X.50.9.741

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *Br. J. Psychol.* 88, 355–383. doi: 10.1111/j.2044-8295.1997.tb02641.x

Michell, J. (2013). Constructs, inferences, and mental measurement. *New Ideas Psychol.* 31, 13–21. doi: 10.1016/j.newideapsych.2011.02.004

Michell, J., and Ernst, C. (1996). The axioms of quantity and the theory of measurement: translated from part I of Otto Hölder's German text die Axiome der Quantität und die Lehre vom mass. *J. Math. Psychol.* 40, 235–252. doi: 10.1006/jmps.1996.0023

Mischel, W., and Shoda, Y. (1994). Personality psychology has tow goals: must it be two fields? *Psychol. Inq.* 5, 156–158. doi: 10.1207/s15327965pli0502_15

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever. *Meas. Interdiscip. Res. Perspect.* 2, 201–218. doi: 10.1207/s15366359mea0204_1

Molenaar, P. C. M. (2008). On the implications of the classical ergodic theorems: analysis of developmental processes has to focus on intra-individual variation. *Dev. Psychobiol.* 50, 60–69. doi: 10.1002/dev.20262

Molenaar, P. C. M., and Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Curr. Dir. Psychol. Sci.* 18, 112–117. doi: 10.1111/j.1467-8721.2009.01619.x

Morin, E. (2008). *On complexity*. Cresskill, NJ: Hampton Press.

Morris, C. W. (1938). *Foundations of the Theory of Signs*. Chicago, IL: University of Chicago Press.

Neuman, Y., Turney, P., and Cohen, Y. (2012). How language enables abstraction: a study in computational cultural psychology. *Integr. Psychol. Behav. Sci.* 46, 129–145. doi: 10.1007/s12124-011-9165-8

Newton, P. E. (2012). Clarifying the consensus definition of validity. *Meas. Interdiscip. Res. Perspect.* 10, 1–29. doi: 10.1080/15366367.2012.669666

Newton, P. E., and Baird, J.-A. (2016). The great validity debate. *Asses. Educ. Principles Policy Pract.* 23, 173–177. doi: 10.1080/0969594X.2016.1172871

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., et al. (2022). Replicability, robustness, and reproducibility in psychological science. *Annu. Rev. Psychol.* 73, 719–748. doi: 10.1146/annurev-psych-020821-114157

Ogden, C. K. (1932). *Bentham's Theory of Fictions*. New York: Harcourt Brace.

Ogden, C. K., and Richards, I. A. (1923). *The Meaning of Meaning: A Study of the Influence of Language Upon Thought and of the Science of Symbolism*. Orlando, FL: Harcourt, Brace & World.

Peirce, C. S. (1958). *Collected Papers of Charles Sanders Peirce*, Vols. *1-6*, C. Hartshorne and P. Weiss (eds.), A. W. Burks (ed.). Cambridge, MA: Harvard University Press. (Vols. 7–8).

Pirnay-Dummer, P., Ifenthaler, D., and Seel, N. M. (2012). "Semantic networks" in *Encyclopedia of the Sciences of Learning*. ed. N. M. Seel (Boston, MA: Springer), 3025–3029.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903. doi: 10.1037/0021-9010.88.5.879

Prigogine, I., and Stengers, I. (1997). *The End of Certainty: Time, Chaos, and the New Laws of Nature*. New York, NY: Free Press.

Quinn, T. J. (2010). *From Artefacts to Atoms: The BIPM and the Search for Ultimate Measurement Standards*. Oxford, UK: Oxford University Press.

Ramlo, S. (2021). Mixed methods research and quantum theory: Q methodology as an exemplar for complementarity. *J. Mixed Methods Res.* 16, 226–241. doi: 10.1177/15586898211019497

Regenwetter, M., and Robinson, M. M. (2017). The construct-behavior gap in behavioral decision research: a challenge beyond replicability. *Psychol. Rev.* 124, 533–550. doi: 10.1037/rev0000067

Richters, J. E. (2021). Incredible utility: the lost causes and causal debris of psychological science. *Basic Appl. Soc. Psychol.* 43, 366–405. doi: 10.1080/01973533.2021.1979003

Robinson, O. C. (2011). The idiographic/nomothetic dichotomy: tracing historical origins of contemporary confusions. *History Philos. Psychol.* 13, 32–39.

Rød, J. (2004). Cartographic signs and arbitrariness. *Cartographica Int. J. Geographic Inf. Geovisualization* 39, 27–36. doi: 10.3138/4462-2125-1312-217T

Roivainen, E. (2013). Frequency of the use of English personality adjectives: implications for personality theory. *J. Res. Pers.* 47, 417–420. doi: 10.1016/j.jrp.2013.04.004

Rosenbaum, P. J., and Valsiner, J. (2011). The un-making of a method: from rating scales to the study of psychological processes. *Theory Psychol.* 21, 47–65. doi: 10.1177/0959354309352913

Rothschuh, K. E. (1963). *Theorie des Organsimus. Bios – Psyche – Pathos (2. Erw. Aufl.)*. München: Urban & Schwarzenberg.

Rotter, J. B. (1954). *Social Learning and Clinical Psychology*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Rudolph, L. (2013) in *Qualitative Mathematics for the Social Sciences. Mathematical Models for Research on Cultural Dynamics*. ed. L. Rudolph (London: Routledge)

Salvatore, S. (2015). *Psychology in Black and White: The Project of a Theory-Driven Science*. Charlotte, NC: Age Information Publishing.

Salvatore, S., Fini, V., Mannarini, T., Valsiner, J., and Veltri, G. A. (2019). *Symbolic Universes in Time of (post)crisis: The Future of European Society*. Cham: Springer.

Salvatore, S., and Valsiner, J. (2010). Between the general and the unique: overcoming the nomothetic versus idiographic opposition. *Theory Psychol.* 20, 817–833. doi: 10.1177/0959354310381156

Sato, T., Hidaka, T., and Fukuda, M. (2009). "Depicting the dynamics of living the life: the trajectory equifinality model" in *Dynamic Process Methodology in the Social and Developmental Sciences*. eds. J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra and N. Chaudhary (New York, NY: Springer), 217–240.

Schwarz, N. (1999). Self-reports: how the questions shape the answers. *Am. Psychol.* 54, 93–105. doi: 10.1037/0003-066X.54.2.93

Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E., and Clark, L. (1991). Rating scales: numeric values may change the meaning of scale labels. *Public Opin. Q.* 55, 570–582. doi: 10.1086/269282

Searle, J. R. (2003). *Rationality in Action*. Cambridge, MA: MIT Press.

Sechrest, L., McKnight, P., and McKnight, K. (1996). Calibration of measures for psychotherapy outcome studies. *Am. Psychol.* 51, 1065–1071. doi: 10.1037/0003-066X.51.10.1065

Semin, G. R., and Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: social cognition and language. *J. Pers. Soc. Psychol.* 54, 558–568. doi: 10.1037/0022-3514.54.4.558

Semin, G. R., and Greenslade, L. (1985). Differential contributions of linguistic factors to memory-based ratings: systematizing the systematic distortion hypothesis. *J. Pers. Soc. Psychol.* 49, 1713–1723. doi: 10.1037/0022-3514.49.6.1713

Shweder, R. A. (1977). Likeness and likelihood in everyday thought: magical thinking in judgments about personality. *Curr. Anthropol.* 18, 637–658. doi: 10.1086/201974

Sim, J., and Mengshoel, A. M. (2022). Metasynthesis: issues of empirical and theoretical context. *Qual. Quant.* doi: 10.1007/s11135-022-01502-w

Simms, L. J., Zelazny, K., Williams, T. F., and Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychol. Assess.* 31, 557–566. doi: 10.1037/pas0000648

Slaney, K. L. (2017). *Validating Psychological Constructs: Historical, Philosophical, and Practical Dimensions*. London, UK: Palgrave Macmillan.

Slaney, K. L., and Garcia, D. A. (2015). Constructing psychological objects: the rhetoric of constructs. *J. Theor. Philos. Psychol.* 35, 244–259. doi: 10.1037/teo0000025

Smedslund, J. (1991). The pseudoempirical in psychology and the case for psychologic. *Psychol. Inq.* 2, 325–338. doi: 10.1207/s15327965pli0204_1

Smedslund, J. (2004). *Dialogues About a New Psychology*. Chagrin Falls, OH: Taos Inst Publications.

Smedslund, J. (2016). Why psychology cannot be an empirical science. *Integr. Psychol. Behav. Sci.* 50, 185–195. doi: 10.1007/s12124-015-9339-x

Smedslund, J. (2021). From statistics to trust: psychology in transition. *New Ideas Psychol.* 61:100848. doi: 10.1016/j.newideapsych.2020.100848

Snider, J. G., and Osgood, C. E. (1969). *Semantic Differential Technique: A Sourcebook*. Chicago: Aldine.

Spearman, C. (1904). General intelligence, objectively determined and measured. *Am. J. Psychol.* 15, 201–293. doi: 10.2307/1412107

Speelman, C. P., and McGann, M. (2020). Statements about the pervasiveness of behavior require data about the pervasiveness of behavior. *Front. Psychol.* 11:594675. doi: 10.3389/fpsyg.2020.594675

Sprung, L., and Sprung, H. (1984). *Grundlagen der Methodologie und Methodik der Psychologie*. Berlin: Deutscher Verlag der Wissenschaften.

Stern, W. (1900). *Über Psychologie der individuellen Differenzen (Ideen zu einer 'differentiellen Psychologie') [On the Psychology of Individual Differences: Ideas about a 'Differential Psychology']*. Leipzig: Barth.

Stern, W. (1911). *Die Differentielle Psychologie in ihren methodischen Grundlagen [Differential Psychology in its Methodological Foundations]*. Leipzig: Barth.

Stevens, S. S. (1935). The operational definition of psychological concepts. *Psychol. Rev.* 42, 517–527. doi: 10.1037/h0056973

Stevens, S. S. (1946). On the theory of scales of measurement. *Science* 103, 667–680.

Strauch, R. E. (1976). Critical look at quantitative methodology. *Policy Anal.* 2, 121–144. Available at: http://www.jstor.org/stable/42784260

Stulík, O. (2022). Recognising and applying equivalent meanings: an example of creating a codebook from a language-game of extremism in the Czech Republic. *Qual. Quant.* doi: 10.1007/s11135-022-01418-5

Szollosi, A., Kellen, D., Navarro, D. J., Shiffrin, R., van Rooij, I., Van Zandt, T., et al. (2020). Is preregistration worthwhile? *Trends Cogn. Sci.* 24, 94–95. doi: 10.1016/j.tics.2019.11.009

Tellegen, A. (1993). Folk concepts and psychological concepts of personality and personality disorder. *Psychol. Inq.* 4, 122–130. doi: 10.1207/s15327965pli0402_12

Thelen, E., and Smith, L. B. (1994). *A Dynamic Systems Approach to the Development of Cognition and Action*. Cambridge, MA: MIT Press.

Thorndike, E. L. (1903). *Notes on Child Study* (*2nd ed.*). New York, NY: Macmillan.

Thurstone, L. L. (1928). Attitudes can be measured. *Am. J. Sociol.* 33, 529–554.

Thurstone, L. L. (1937). Psychology as a quantitative rational science. *Science* 85, 227–232. doi: 10.1126/science.85.2201.227

Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.

Toomela, A. (2003). Relationships between personality structure, structure of word meaning, and cognitive ability: a study of cultural mechanisms of personality. *J. Pers. Soc. Psychol.* 85, 723–735. doi: 10.1037/0022-3514.85.4.723

Toomela, A. (2008). Variables in psychology: a critique of quantitative psychology. *Integr. Psychol. Behav. Sci.* 42, 245–265. doi: 10.1007/s12124-008-9059-6

Toomela, A. (2009). "How methodology became a toolbox—and how it escapes from that box" in *Dynamic Process Methodology in the Social and Developmental Sciences*. eds. J. Valsiner, P. C. M. Molenaar, M. C. D. P. Lyra and N. Chaudhary (New York, NY: Springer), 45–66. doi: 10.1007/978-0-387-95922-1_3

Toomela, A. (2018). *The Psychology of Scientific Inquiry*. Cham: Springer International Publishing.

Toomela, A. (2021). *Culture, Speech and Myself*. Harjumaa, Estonia: Porcos ante Margaritas.

Toomela, A., and Valsiner, J. (2010). *Methodological Thinking in Psychology: 60 Years Gone Astray?* Charlotte, NC: Information Age Publishing.

Tourangeau, R., Rips, L. J., and Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.

Trendler, G. (2009). Measurement theory, psychology and the revolution that cannot happen. *Theory Psychol.* 19, 579–599. doi: 10.1177/0959354309341926

Trendler, G. (2019). Conjoint measurement undone. *Theory Psychol.* 29, 100–128. doi: 10.1177/0959354318788729

Trendler, G. (2022). The incoherence of Rasch measurement: a critical comparison between measurement in psychology and physics. *Personal. Individ. Differ.* 189:111408. doi: 10.1016/j.paid.2021.111408

Uher, J. (2011). Individual behavioral phenotypes: an integrative meta-theoretical framework. Why behavioral syndromes are not analogs of personality. *Dev. Psychobiol.* 53, 521–548. doi: 10.1002/dev.20544

Uher, J. (2013). Personality psychology: lexical approaches, assessment methods, and trait concepts reveal only half of the story-why it is time for a paradigm shift. *Integr. Psychol. Behav. Sci.* 47, 1–55. doi: 10.1007/s12124-013-9230-6

Uher, J. (2015a). "Agency enabled by the psyche: explorations using the transdisciplinary philosophy-of-science paradigm for research on individuals" in *Constraints of Agency: Explorations of Theory in Everyday Life. Annals of Theoretical Psychology*. eds. C. W. Gruber, M. G. Clark, S. H. Klempe and J. Valsiner, vol. *12* (New York: Springer International Publishing), 177–228. doi: 10.1007/978-3-319-10130-9_13

Uher, J. (2015b). "Comparing individuals within and across situations, groups and species: metatheoretical and methodological foundations demonstrated in primate behaviour" in *Comparative Neuropsychology and Brain Imaging (Vol. 2), Series Neuropsychology: An Interdisciplinary Approach*. eds. D. Emmans and A. Laihinen (Berlin: Lit Verlag), 223–284.

Uher, J. (2015c). Conceiving "personality": Psychologist's challenges and basic fundamentals of the transdisciplinary philosophy-of-science paradigm for research on individuals. *Integr. Psychol. Behav. Sci.* 49, 398–458. doi: 10.1007/s12124-014-9283-1

Uher, J. (2015d). Developing "personality" taxonomies: metatheoretical and methodological rationales underlying selection approaches, methods of data generation and reduction principles. *Integr. Psychol. Behav. Sci.* 49, 531–589. doi: 10.1007/s12124-014-9280-4

Uher, J. (2015e). Interpreting "personality" taxonomies: why previous models cannot capture individual-specific experiencing, behaviour, functioning and development. Major taxonomic tasks still lay ahead. *Integr. Psychol. Behav. Sci.* 49, 600–655. doi: 10.1007/s12124-014-9281-3

Uher, J. (2016a). "Exploring the workings of the psyche: metatheoretical and methodological foundations" in *Psychology as the Science of Human Being: The Yokohama Manifesto*. eds. J. Valsiner, G. Marsico, N. Chaudhary, T. Sato and V.

Dazzani (New York: Springer International Publishing), 299–324. doi: 10.1007/978-3-319-21094-0_18

Uher, J. (2016b). What is behaviour? And (when) is language behaviour? A metatheoretical definition. *J. Theory Soc. Behav.* 46, 475–501. doi: 10.1111/jtsb.12104

Uher, J. (2018a). Quantitative data from rating scales: an epistemological and methodological enquiry. *Front. Psychol.* 9:2599. doi: 10.3389/fpsyg.2018.02599

Uher, J. (2018b). Taxonomic models of individual differences: a guide to transdisciplinary approaches. *Philos. Trans. R. Soc. B* 373:20170171. doi: 10.1098/rstb.2017.0171

Uher, J. (2018c). "The Transdisciplinary Philosophy-of-Science Paradigm for Research on Individuals: foundations for the science of personality and individual differences" in *The SAGE Handbook of Personality and Individual Differences: Volume I: The Science of Personality and Individual Differences*. eds. V. Zeigler-Hill and T. K. Shackelford (London, UK: SAGE), 84–109.

Uher, J. (2019). Data generation methods across the empirical sciences: differences in the study phenomena's accessibility and the processes of data encoding. *Qual. Quantity Int. J. Method.* 53, 221–246. doi: 10.1007/s11135-018-0744-3

Uher, J. (2020a). Human uniqueness explored from the uniquely human perspective: epistemological and methodological challenges. *J. Theory Soc. Behav.* 50, 20–24. doi: 10.1111/jtsb.12232

Uher, J. (2020b). Measurement in metrology, psychology and social sciences: data generation traceability and numerical traceability as basic methodological principles applicable across sciences. *Qual. Quantity Int. J. Method.* 54, 975–1004. doi: 10.1007/s11135-020-00970-2

Uher, J. (2021a). Problematic research practices in psychology: misconceptions about data collection entail serious fallacies in data analysis. *Theory Psychol.* 31, 411–416. doi: 10.1177/09593543211014963

Uher, J. (2021b). Psychology's status as a science: peculiarities and intrinsic challenges. Moving beyond its current deadlock towards conceptual integration. *Integr. Psychol. Behav. Sci.* 55, 212–224. doi: 10.1007/s12124-020-09545-0

Uher, J. (2021c). Psychometrics is not measurement: Unraveling a fundamental misconception in quantitative psychology and the complex network of its underlying fallacies. *J. Theor. Philos. Psychol.* 41, 58–84. doi: 10.1037/teo0000176

Uher, J. (2021d). Quantitative psychology under scrutiny: measurement requires not result-dependent but traceable data generation. *Personal. Individ. Differ.* 170:110205. doi: 10.1016/j.paid.2020.110205

Uher, J. (2022). Functions of units, scales and quantitative data: fundamental differences in numerical traceability between sciences. Quality & Quantity. *Int. J. Met.* 56, 2519–2548. doi: 10.1007/s11135-021-01215-6

Uher, J., Addessi, E., and Visalberghi, E. (2013a). Contextualised behavioural measurements of personality differences obtained in behavioural tests and social observations in adult capuchin monkeys (Cebus apella). *J. Res. Pers.* 47, 427–444. doi: 10.1016/j.jrp.2013.01.013

Uher, J., and Visalberghi, E. (2016). Observations versus assessments of personality: a five-method multi-species study reveals numerous biases in ratings and methodological limitations of standardised assessments. *J. Res. Pers.* 61, 61–79. doi: 10.1016/j.jrp.2016.02.003

Uher, J., Werner, C. S., and Gosselt, K. (2013b). From observations of individual behaviour to social representations of personality: developmental pathways, attribution biases, and limitations of questionnaire methods. *J. Res. Pers.* 47, 647–667. doi: 10.1016/j.jrp.2013.03.006

Valsiner, J. (1997). *Culture and the Development of Children's Actions: A Theory of Human Development* (*2nd ed.*). Chichester: John Wiley & Sons.

Valsiner, J. (1998). *The Guided Mind: A Sociogenetic Approach to Personality*. Cambridge, MA: Harvard University Press.

Valsiner, J. (2001). Process structure of semiotic mediation in human development. *Hum. Dev.* 44, 84–97. doi: 10.1159/000057048

Valsiner, J. (2012). *A Guided Science: History of Psychology in the Mirror of its Making*. New Brunswick, NJ: Transaction Publishers.

Valsiner, J. (2014). Needed for cultural psychology: methodology in a new key. *Cult. Psychol.* 20, 3–30. doi: 10.1177/1354067X13515941

Valsiner, J. (2017a). Beyond the three snares: implications of James' 'psychologist's fallacy' for 21st century science. *New Ideas Psychol.* 46, 26–30. doi: 10.1016/j.newideapsych.2016.05.001

Valsiner, J. (2017b). *From Methodology to Methods in Human Psychology*. Cham: Springer International Publishing.

Valsiner, J. (2021). *General Human Psychology*. Springer Nature Switzerland.

Valsiner, J., Diriwächter, R., and Sauck, C. (2005). "Diversity in Unity: standard questions and nonstandard interpretations" in *Science and Medicine in Dialogue*. eds. R. Biface, J. Laird, K. Noller and J. Valsiner (Westport, CT: Praeger-Greenwood), 289–307.

van Geert, P. (2011). The contribution of complex dynamic systems to development. *Child Dev. Perspect.* 5, 273–278. doi: 10.1111/j.1750-8606.2011.00197.x

Van Geert, P., and Van Dijk, M. (2002). Focus on variability: new tools to study intra-individual variability in developmental data. *Infant Behav. Dev.* 25, 340–374. doi: 10.1016/S0163-6383(02)00140-6

Vanrullen, R. (2011). Four common conceptual fallacies in mapping the time course of recognition. *Front. Psychol.* 2:365. doi: 10.3389/fpsyg.2011.00365

Vautier, S., Lacot, É., and Veldhuis, M. (2014). Puzzle-solving in psychology: the neo-Galtonian vs. nomothetic research focuses. *New Ideas Psychol.* 33, 46–53. doi: 10.1016/j.newideapsych.2013.10.002

Veraksa, N., Basseches, M., and Brandão, A. (2022). Dialectical thinking: a proposed foundation for a post-modern psychology. *Front. Psychol.* 13:710815. doi: 10.3389/fpsyg.2022.710815

Vessonen, E. (2017). Psychometrics versus representational theory of measurement. *Philos. Soc. Sci.* 47, 330–350. doi: 10.1177/0048393117705299

von Bertalanffy, L. (1937). *Das Gefüge des Lebens*. Leipzig: Teubner.

von Eye, A., and Bogat, G. A. (2006). Person-oriented and variable-oriented research: concepts, results, and development. *Merrill-Palmer Q.* 52, 390–420. doi: 10.1353/mpq.2006.0032

von Uexküll, J. (1909). *Umwelt und Innenwelt der Tiere*. Berlin: Springer.

Vygotsky, L. S. (1962). *Thought and Language*. Cambridge, MA: MIT Press.

Wagoner, B., and Valsiner, J. (2005). "Rating tasks in psychology: from a static ontology to a dialogical synthesis of meaning" in *Contemporary Theorizing in Psychology: Global Perspectives*. eds. A. Gülerce, I. Hofmeister, G. Saunders and J. Kaye (Toronto, Canada: Captus), 197–213.

Walach, H. (2013). *Psychologie: Wissenschaftstheorie, Philosophische Grundlagen Und Geschichte* (*3. Aufl.*). Stuttgart: Kohlhammer.

Walach, H., and Römer, H. (2011). Complementarity is a useful concept for consciousness studies: a reminder. *Neuroendocrinol. Lett.* 21, 221–232.

Wallach, L. (1971). Implications of recent work in philosophy of science for the role of operational definition in psychology. *Psychol. Rep.* 28, 583–608. doi: 10.2466/pr0.1971.28.2.583

Weber, M. (1949) in *On the Methodology of the Social Sciences*. eds. E. Shils and H. Finch (New York: Free Press)

Westerman, M. A. (2006). Quantitative research as an interpretive enterprise: the mostly unacknowledged role of interpretation in research efforts and suggestions. *New Ideas Psychol.* 24, 189–211. doi: 10.1016/j.newideapsych.2006.09.004

Westerman, M. A. (2014). Examining arguments against quantitative research: "case studies" illustrating the challenge of finding a sound philosophical basis for a human sciences approach to psychology. *New Ideas Psychol.* 32, 42–58. doi: 10.1016/j.newideapsych.2013.08.002

Whitehead, A. N. (1929). *Process and Reality*. New York, NY: Harper.

Wiggins, J. S. (1979). A psychological taxonomy of trait-descriptive terms: the interpersonal domain. *J. Pers. Soc. Psychol.* 37, 395–412. doi: 10.1037/0022-3514.37.3.395

Windelband, W. (1904/1998). History and natural science. *Theory Psychol.* 8, 5–22. doi: 10.1177/0959354398081001

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London, UK: Kagan Paul.

Wittmann, W. W. (1988). "Multivariate reliability theory: principles of symmetry and successful validation strategies" in *Handbook of Multivariate Experimental Psychology*. eds. J. R. Nesselroade and R. B. Cattell. *2nd* ed (New York, NY: Plenum Press), 505–560. doi: 10.1007/978-1-4613-0893-5_16

Wood, A. M., Brown, G. D. A., Maltby, J., and Watkinson, P. (2012). How are personality judgments made? A cognitive model of reference group effects, personality scale responses, and behavioral reactions. *J. Pers.* 80, 1275–1311. doi: 10.1111/j.1467-6494.2012.00763.x

Wundt, W. (1896). *Grundriss der Psychologie [Outlines of Psychology]*. Stuttgart: Körner

Wundt, W. (1902-1903). *Grundzüge der Physiologischen Psychologie Band 1–3* (*5th*). Leipzig: Engelmann

Wundt, W. (1907). *Logik der exakten Wissenschaften [Logic of the Exact Sciences]. II* Stuttgart: Enke. (3. umgearb. Aufl.)

Yarkoni, T. (2022). The generalizability crisis. *Behav. Brain Sci.* 45:e1. doi: 10.1017/S0140525X20001685

Young, B., Boccaccini, M. T., Conroy, M. A., and Lawson, K. (2007). Four practical and conceptual assessment issues that evaluators should address in capital case mental retardation evaluations. *Prof. Psychol. Res. Pract.* 38, 169–178. doi: 10.1037/0735-7028.38.2.169

Zagaria, A., Ando, A., and Zennaro, A. (2020). Psychology: a giant with feet of clay. *Integr. Psychol. Behav. Sci.* 54, 521–562. doi: 10.1007/s12124-020-09524-5

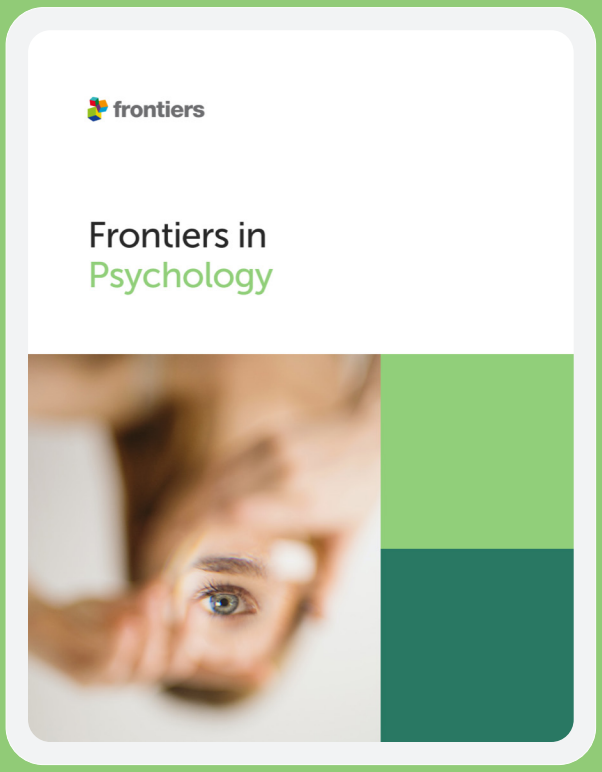
Frontiers in Psychology

Paving the way for a greater understanding of human behavior

The most cited journal in its field, exploring psychological sciences - from clinical research to cognitive science, from imaging studies to human factors, and from animal cognition to social psychology.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact